

C457  
D

372310

段新生 著

# 证据理论与决策、 人工智能

中国人民大学出版社

• 中青年学者文库



(京)新登字156号

**证据理论与决策、人工智能**

段新生 著

\*

中国人民大学出版社出版发行

(北京海淀路39号 邮码 100072)

北京市丰台区丰华印刷厂印刷

(北京鼓楼西大石桥胡同61号)

新华书店经销

\*

开本: 787×1092毫米32开 印张: 7 插页2

1993年3月第1版 1993年3月第1次印刷

字数: 148 000 册数: 1-2 000

\*

ISBN 7-300-01492-5

G·153 定价: 4.30元

## 内 容 简 介

本书介绍了证据理论的基本内容，叙述了作者在理论上所做的部分工作，探讨了证据理论的应用问题。

全书共分十三章，前七章综述证据理论的基本内容，第八章是作者在理论上所做的部分工作，其中包括合成法则的公理化研究，多层证据的处理以及无限框架上证据处理的近似转化等。后几章是作者对证据理论的应用问题的探讨。其中第九章提出了利用证据理论作决策的一般方法，第十章到第十二章利用这种方法研究了决策分析中主观概率的估计问题、预测问题、指标体系的确定问题、领导班子的选取问题以及医疗诊断问题等。最后一章探讨了证据理论在人工智能以及专家系统中的应用。

本书力求简洁明了，通俗易懂，不过分追求严密性，有一定的探索性。不仅适合于对证据理论研究感兴趣的数学工作者以及计算机工作者，更适合于想在自己的研究领域中寻找新工具的社会科学工作者。另外，对于想扩大自己知识面的文、理科学生也是一本合适的读物

## 《中青年学者文库》序

也许我们有点莽撞，在丛书如林的今天，在出版事业步履维艰的今天，还要推出这样一套并非出自名家手笔的《中青年学者文库》，有竞争力吗？前景如何？

我们深知我们力量微薄。但是，坦率地说，我们仍要参与竞争，我们渴望参与竞争。是竞争的时代激励了我们的竞争意识，我们希望能竞争中锻炼队伍，能在竞争中尽快走向成熟。

催促学术队伍的更新和新人的成长，是我们编辑这套《文库》最基本的出发点。我们的事业急切地呼唤着新人的崛起；锐意进取的年轻人切盼着施展才华的开阔地。正是这种双向的需求，激发我们建全起这片小小的人材“特区”，进入这方特区，凭借的不是学历、资历，也不是职称或职务，而是求新的意识和苦心的研究。我们希望我们的作者能够既有劈山开路的气魄，又有滴水穿石的韧性。

我们的作者，有的已崭露头角，有的则刚刚踏上论坛。他们的著作，在学长们面前也许会显得稚拙，甚至偏激。但是，对真理的执著追求却是我们共同的愿望。不趋时，不唯上，不因循守旧，不故弄虚玄，坚持以马克思主义为指导，紧密结合我国国情，广泛吸收世界当代文明中一切有价值的东西，竭力回答社会变革中提出的重大问题，勇于探索学术发展中出现的困难问题，敢于触及社会舆论关注的敏感问

题，——遵循这样的原则，经过不懈的努力，相信我们能有所作为。

草创的事业，期待着各方的支持。高质量的 赐稿是支持，建设性的意见和有份量的批评也是极好的支持。我们确信能够赢得支持，并在热情的支持中不断前进。

《中青年学者文库》编委会

## 写在前面

段新生同志的《证据理论与决策、人工智能》一书就要出版了，我感到非常高兴。

证据理论是1976年由美国学者G. Shafer创立的。证据理论创立以来，理论上取得了很大的发展，应用上也取得了很大的成果。

在我国，证据理论的研究还很薄弱。目前，有关证据理论的内容只能在杂志上或某些书中见到一些零星的介绍。

《证据理论与决策、人工智能》一书系统、扼要地介绍了证据理论的基本内容，这对于将证据理论全面地介绍到国内无疑是非常有意义的。

证据理论是实用性很强的学科。在段新生同志所从事的证据理论的研究中，很大一部分就是证据理论的应用。作者首次把证据理论引入决策科学中，并给出了用信度函数做判决的一般方法；在人工智能的各个领域中，作者应用证据理论也取得了许多很好的成果（例如软推理）。所有这些无疑将给实际工作者一定的启迪。

当然，本书也存在着许多不足之处，但是有不足就要研究，就要探索。相信本书的出版会吸引更多的人加入到有关

证据理论的研究行列中来。

**寿玉亭 1980.8**

## 前 言

1976年, G.Shafer出版了“A Mathematical Theory of Evidence”一书, 标志着证据理论的诞生。证据理论从诞生到现在虽然仅仅经过十几年的时间, 但是理论上却取得许多丰硕的成果, 得到了国际学术界的普遍重视。

本书第八章是作者在理论上所作的几点探讨。其中§1谈了作者对证据合成的看法, 提出了证据合成的公理化方法; §2提出了多层证据的概念及一些处理方法; §3应用粗糙集理论研究了无穷框架上的证据处理向有穷框架上的证据处理的转化问题, 得到了一种无穷框架上证据处理的近似方法。

证据理论一诞生, 专家系统的许多建造者就注意到了这种理论。由于证据理论可以作为研究某种不确定性推理的理论, 因而专家系统的建造者就试图利用这种理论处理专家系统中大量存在的不确定性问题。利用证据理论建造专家系统取得了一定的成果, 但是还有许多不完善的地方。这种研究尚在进行之中。

但是, 自证据理论诞生以来, 也只有专家系统的研究者和建造者们在应用证据理论。证据理论在其他领域中的应用却几乎是一个无人问津的问题。



我们注意到，证据理论是一种关于判决的理论，因此在所有包含有判决的领域中都可以尝试应用证据理论。

决策领域包含有大量的判决问题，所以，作者认为，证据理论的一个最适合应用的领域就是决策。本书第九章探讨了利用证据理论作决策的一般方法，第十章至第十二章探讨了这种方法在某些具体领域中的应用。其中第十章是在决策分析领域中的应用。在用决策分析的方法作决策时，常常遇到系统生存状态的主观概率的估计问题，利用第九章的方法可以给出一种综合各专家经验和知识的专家咨询估计法，同时在该章还得到了另外一种决策分析的方法——M决策法。第十一章是在预测领域中的应用。本章利用第九章方法得到了一种预测的方法——信度预测法。信度预测法是特尔菲预测法的推广，但无论在咨询的方式上还是在专家意见的表述及处理上都与特尔菲预测法有本质的不同。同时本章还得到了一种先预测状态再进行决策的一种决策分析的方法。第十二章是在其他领域中的应用。这些领域诸如指标体系的确定、领导班子的选取以及医疗诊断等。证据理论在这些领域中的应用与以上应用有一些区别。证据理论不能直接用于这些问题的解决，但是使用一些简单的技巧，第九章的方法仍然可以在这些领域中得到应用。

社会科学中包含有大量的决策问题。因此，作者认为，利用证据理论作决策的一般方法可以在社会科学的许多领域中得到应用（并不仅仅是本书所涉及的几个领域），可以作为社会科学定量研究的一种工具。

另外，专家系统的研究者和建造者对证据理论的应用仅仅局限于人工智能的专家系统领域，人工智能的其他领域却

无人问津。本书最后一章探讨了证据理论在推理（包括软推理，归纳推理以及非单调推理）以及模式识别中的应用。在这一章最后一节还研究了专家系统中各种不确定性的存在，探讨了这些不确定性的处理，提出了一种建造以信度函数为基础而不是以产生式规则为基础的专家系统的思想。

最后感谢袁萌老师，魏晴宇老师，张金马老师的帮助，感谢陈禹老师对本书出版所作的大量工作，感谢寿玉亭老师在百忙中阅读了本书的初稿并提出了许多宝贵的意见。对本书缺点及错误，恳请读者批评指正。

段新生 1989.6.2

# 目 录

第一章	序言	1
§1	概率的三种解释	1
§2	构造性解释	3
§3	信度所遵循的原则	5
§4	证据理论的发展	6
§5	证据理论的应用	8
第二章	信度函数	13
§1	识别框架	13
§2	基本可信度分配与信度函数	14
§3	众信度函数与似真度函数	19
§4	$m$ 、 $Bel$ 、 $pl$ 、 $Q$ 的几何意义	23
§5	贝叶斯信度函数	24
第三章	Dempster合成法则	26
§1	两个信度函数的合成	26
§2	多个信度函数的合成	30
§3	Dempster合成法则的众信度函数形式	32

§4	冲突权重	33
§5	条件信度函数	34
第四章	框架的转化	36
§1	粗化与细分	36
§2	相容框架族	40
§3	相容的信度函数	43
§4	相容框架的独立性	44
§5	框架的收缩与扩张	45
第五章	信度函数的分类及证据的权重	47
§1	简单支持函数	47
§2	可分离支持函数	50
§3	支持函数	52
§4	证据的权重	54
§5	准支持函数	62
§6	信度函数的分类	63
第六章	一致支持函数	65
§1	一致支持函数	65
§2	外形函数	67
§3	证据的一致性	69
§4	推理证据	71
§5	一致支持函数与可能性分布	72
第七章	统计证据	76

§1	引言	76
§2	单个观察值下的统计证据	77
§3	统计证据的权重	78
§4	多个观察值下的统计证据	81
§5	统计证据的折扣	82
§6	在框架的相容变换下支持函数的变化	83
<b>第八章</b>	<b>理论上的进一步探讨</b>	<b>86</b>
§1	证据合成的公理化研究	86
§2	多层证据处理	97
§3	无限框架上证据的近似处理	102
<b>第九章</b>	<b>利用Shafer证据理论作决策的一般方法</b>	<b>114</b>
§1	问题的提出	114
§2	用“最少点”原则求真值	115
§3	利用Shafer证据理论作决策的一般算法	118
§4	一些问题的讨论	124
<b>第十章</b>	<b>决策分析</b>	<b>126</b>
§1	主观概率的传统估计	127
§2	专家咨询法	129
§3	Dempster合成法则的递归形式 及主观概率的修正	131
§4	M决策法	133
<b>第十一章</b>	<b>预测</b>	<b>140</b>





# 第一章 序 言

## §1 概率的三种解释

在Shafer的证据理论出现之前(1976年以前),概率(probability)的解释可以概括为三种:客观解释,个人主义解释以及必要性解释(Objectivistic, Personalistic, Necessary)。<sup>1</sup>根据客观解释,概率描述了一个可以重复出现的事件的客观事实,用该事件发生的频率当试验次数趋于无穷时的极限来刻画。根据个人主义解释,概率是某个个人的偏好(Opinion),它可以根据某个人在赌博中或其他带有不确定性结果的事件中所表现出来的行为来推算。根据必要性解释,概率是测量一个命题推出另一个命题的程度的量,这个量由两个命题之间的逻辑关系完全决定,与人类的偏好无任何关系。

客观解释又叫频率解释。根据这种解释,要求出一个事件的概率必须要求出该事件重复出现的频率。比如说,许多次地掷一个骰子,出现1的次数占总投掷次数的 $1/6$ ,而且次数越多,越接近 $1/6$ ,那么我们就说掷骰子出现1的概率为 $1/6$ 。除此而外,如果一个事件是不可重复的,那么这个事件也就无概率可言。举例来说,要问“明天下雨”的概率是

多少，或“一个刚出生的婴儿能活到70岁”的概率是多少，在频率主义者看来是无意义的。原因就是，“明天下雨”以及“刚出生的婴儿能活到70岁”这样的事件是不可重复的，因而没法谈论它们的概率。数学概率论以及数理统计所采用的概率就是在一种广泛频率解释下的概率。因此大多数数理统计学家都是频率主义者。

个人主义解释又称为主观解释或贝叶斯解释。在这种解释看来，概率与某个人的行为相联系，由某人在进行赌博或某种带有不确定性结果的活动中所表现出来的行为来反映。比如，一个人在打赌时经常为“白马取胜”下较大的赌注，那么在他看来“白马取胜”的概率就很大。而且他下的赌注越大，那么概率就越大。因此个人主义的概率反映了个人的一种偏好，是个人的主观意愿作用的结果。说“明天下雨”的概率是0.8，即是说该人对“明天下雨”的主观推断是0.8。主观解释是当今最流行的概率解释，L.J.Savage 是其典型代表，在他写的“*The Foundations of statistics*” (1954) 一书中，他的观点得到了充分的表述，他在书中导出的基于这种解释的决策分析与效用理论在当今人们的决策活动中起到了非常大的作用。

值得指出的是，现在许多频率主义者也经常采用主观主义的观点。例如，在数学概率论中有贝叶斯条件法则。尽管频率主义者根本不承认先验概率的存在，但是要应用贝叶斯条件法则就必须首先给出先验概率，而先验概率的给出在没有大量统计资料可利用的情况下必须依靠人的主观估计。

必要性解释又叫逻辑主义解释。逻辑主义解释把概率看成是命题与命题之间联系程度的度量。这种联系程度是纯客



观的，与人的作用毫无关系。逻辑主义解释的代表人物是 M. Keynes。Keynes 把概率理论看作一种逻辑学，认为它的对象是人的推理过程。概率理论研究的推理是可以合理地赋予某个权重的推理。如果说演绎推理由前提必然可以得到结论，那么概率推理由前提可以在一定程度上得到结论。从这个意义上讲，演绎推理是概率推理的特例，而概率推理又是归纳推理的一种。逻辑主义解释使得归纳逻辑得到了很大发展。

## §2 构造性解释

Shafer 证据理论给概率一种新的解释。

Shafer 指出：以上三种解释都没涉及概率推断的构造性特征。频率主义解释以及逻辑主义解释赋予概率一种客观属性，概率的得到与人类活动毫无关系。贝叶斯解释把概率解释成人的偏好或主观意愿的度量，但是它既没强调也没要求概率如何构造。<sup>[4]</sup>

不管是哪种解释，概率的得到都可以看成是在为一个命题赋真值，只不过该真值并非非真即假或非假即真，而是可以取  $[0, 1]$  之间的所有值，换句话说该命题为真的程度并非 1 或 0，而是介于 0 与 1 之间的数。我们求某个命题的概率也即确定它为真的程度。例如，我们说“明天下雨”的概率是 0.5，就是说命题“明天下雨”为真的程度为 0.5。

因此为命题赋予真值在贝叶斯解释看来，完全是人主观判决的结果，与证据毫无关系，尽管有时也谈到证据，但是根本不强调证据，即贝叶斯解释片面强调人的判决作用，忽

视证据的作用。为命题赋予真值在逻辑主义解释看来完全是证据作用的结果，是证据（作为前提）与该命题逻辑关系的体现，是纯客观的。频率解释也将概率归结为一个事件的客观属性，是由该事件发生的频率（作为证据）完全决定的。因此逻辑主义解释与频率主义解释片面强调证据的作用，忽视人的判决作用。

Shafer指出：对于概率推断的理解，我们不仅要强调证据的客观性而且也要强调证据估计的主观性。数字化的概率并没有独立于人类判决的客观属性，在人思考之前也不会在人的头脑中存在。但是我们人可以在客观证据的基础上构造出这样一个数字化的概率来。<sup>[8]</sup>

在构造该概率的时候，我们可以分析证据。一开始，对证据的理解可能是含混的，得到的概率也可能是非数字化的，无任何结构的，但随着这种含混的、混乱的感觉的逐渐深入，最后，他可以确定出一个数字用来表示他对该命题为真的信任程度(degree of belief)。

综上所述，根据构造性解释(Constructive interpretation)，概率是某人在证据的基础上构造出的他对一命题真的信任程度，简称为信度。

例如，某人说“明天下雨”的概率是0.5，即是说，根据某些迹象（证据），比如，天气闷热，喘气困难等，经过分析得出的他对“明天下雨”为真的信任程度。

本书第二章到第七章所介绍的就是这种在构造性解释下的概率理论，因此是另一种贝叶斯主观概率理论。

为了强调这种理论对概率的不同理解，〔1〕中出现的概率(probability)一词不译为概率而译为信度或可信度等，

比如：basic probability assignement译为基本可信度分配，probability judgement译为可信度推断等。而这种理论我们也不称为是一种概率理论（由下节可以看出，这种理论确实不能称为概率理论），而称为证据的数学理论或简称为证据理论，有时为突出信度也称为信度理论等。

在第二章，读者还能看到概率的构造性解释的更详细的说明。

### §3 信度所遵循的原则

§1的三种解释下的概率都必须满足可加性，即

$$\forall A, B \subset \Theta \quad A \cap B = \phi$$

$$P(A \cup B) = P(A) + P(B)$$

其中， $\Theta$ 为样本空间， $A, B$ 为 $\Theta$ 中的事件。

这个特性是以往概率理论所普遍遵循的一个原则。

关于可加性，贝叶斯主义者声称，这是人类概率所必须遵循的普遍原则。如果一个人产生了不满足可加性的概率，那么他将是无理智的，荒谬的。可见，贝叶斯主义者排斥其他的概率理论。

根据可加性，如果我们相信一个命题为真的程度为 $s$ ，那么我们就必须以 $1-s$ 的程度去相信该命题的反。在许多情况下，这是不合理的。举例来说，对“地球以外存在着生命”和“地球以外不存在生命”这一对命题来说，在目前科学水平或我们目前所拥有的知识结构（证据）下，我们既不相信前者，又不敢相信后者，即我们对前者的信度很小，对后者的信度也很小，因此两者之和根本不可能等于1。

因此，对于信度，Shafer舍弃了这样一个原则，而用一种称为半可加性的原则来代替。

对于这种半可加性原则，Shafer指出：我们并不要求一个人所产生的信度必须遵循这个原则，也不去说如果一个人产生了遵循其他原则的信度就是无理智的，荒谬的。<sup>[11]</sup>那么为什么采用这个原则而不采用其他原则呢？Shafer回答说：回答这个问题就相当于回答为什么一种语言要采用这种语言所遵循的语法结构一样是毫无意义的。不过要解释的话也可以从整个理论的内部作些解释。由于这种半可加性与理论的其他方面是协调的，比如说，只有满足这个原则的函数才能用Dempster合成法则（见第三章）进行合成，而且在某个具体例子（码的传输）中可以产生遵循这种原则的函数，因此使用半可加性是合理的。<sup>[12]</sup>但是，Shafer又接着指出：这种解释仅仅是一种理论内部的解释，不能排除遵循其他原则的信度函数的存在；而且也不能担保，遵循其他原则的信度函数的理论就不比他的理论更有用，更优越。

具体的半可加性原则请见第二章。

#### §4 证据理论的发展

1976年，Shafer出版了《证据的数学理论》一书。该书的出版标志着证据理论的诞生。十几年来，许多作者又为证据理论的发展作出了许多贡献。现在证据理论已经取得了许多丰硕的成果，得到了国际学术界的承认。

为证据理论作出重大贡献的第一个人物是A.P.Dempster。Dempster 1967年曾经在[14]中给出过上、下概率的概

念，第一次明确给出了不满足可加性的概率。1968年，他又在[15]中探讨了统计推理的一般化问题，而且针对统计问题给出了两批证据（即两个独立的信息源）合成的原则。Shafer证据理论就是在Dempster工作的基础上产生的。在Shafer证据理论中，最重要的合成法则——Dempster合成法则也是Dempster在研究统计问题时首先给出的，Shafer只不过把它推广到更加一般的情形。为了纪念Dempster对该理论的贡献，有人也称证据理论为Dempster—Shafer理论。

D. Dubois和H. Prade为证据理论的发展也作出了贡献。在[12]中，他们从数学形式上研究了信度函数，得出了信度函数是一模糊测度的结论。另外，在信度理论与可能性理论的联系中也得出了许多有益的结果。在[13]中，他们又以一种所谓的集论观点分析、研究了证据，得到了诸如证据的并、证据的交、证据的补以及证据的包含等概念。现在他们仍然在为证据理论的发展辛勤工作着。

模糊数学的创始人L. A. Zadeh也为证据理论的发展作出了大量的贡献。他在[16]中谈了他对Dempster合成法则的看法，尤其指出归一化常数 $K$ 加得有点勉强，建议将 $K$ 去掉，引入假设 $m(\phi) > 0$ ，而 $m(\phi) > 0$ 意味着我们的真值可以在框架 $\Theta$ 之外。另外，在其他的文章中，他也曾经研究过证据理论与可能性理论的关系。

P. Smits是另一位对证据理论作出过贡献的人。他在[10]中将信度函数推广到框架 $\Theta$ 的所有模糊子集上，得出了许多有益的结果。在[11]中，他又谈了将模糊集和信度函数用于医疗诊断的看法。

还有许多其他作者都对证据理论的发展作出过贡献。在

此不一一介绍。

另外，本书限于篇幅，对以上作者的工作也不能作详细的阐述。因此书中涉及到的证据理论，基本上是Shafer的基本的证据理论。

本书第八章是作者在理论上所作的几点探讨。其中§1谈了作者对Dempster合成法则的看法；通过引入定义不完全的代数系统的概念，提出了一种研究证据合成的公理化方法。§2给出了多层证据处理的概念，并得到了一些多层证据处理的方法。§3针对Shafer无限框架上信度函数的定义，应用Z.Pawlak1982年创立的粗糙集理论<sup>[9]</sup>给出了无限框架上的证据处理向有限框架上的证据处理的近似转化，从而得到了一种比较可行的无限框架上的证据处理的方法。

## §5 证据理论的应用

证据理论究竟如何来应用？在什么领域中可以应用？本书第九章到第十三章就是作者对这两个问题所作的回答。

就作者看来，证据理论首先可以用于决策领域。

在§2概率的构造性解释中曾经说过，证据理论强调依据证据为一个命题赋予真值。为一个命题赋予一个真值本身可以看成是一种决策模型，因此证据理论可以看成是根据证据做决策的理论。从这个意义上讲，证据理论理所当然可以用于决策问题的解决。

证据理论为我们提供了证据作用如何描述的方法。对一个判决问题，设我们所能认识到的可能结果用集合 $\Theta$ 表示， $\Theta$ 称为我们的识别框架。如果有一批针对着该框架的证据的话，

那么依据证据理论，我们就可以在框架 $\Theta$ 上产生一个信度函数。信度函数是该批证据作用的结果。它反映了根据该批证据对框架 $\Theta$ 可识别的各个命题所赋予的真值。如果有多批证据同时作用于框架 $\Theta$ ，那么我们可以用Dempster合成法则合成依据每批证据所得到的信度函数，该信度函数可以表示以上各批证据的联合作用。因此在证据理论看来，依据证据所作出的判决结果就是信度函数。有了信度函数，我们可以回答判决的真值在某集合中的可能性有多大的问题。但是依据这样的结果，我们却不能回答判决的结果究竟是什么的问题。可是，实际的决策却常常要求我们对结果作出明确的回答。因此利用证据理论作决策时还必须解决依据信度函数如何得出明确的判决结果的问题。本书第九章就讨论了这样的问题。在第九章我们得到了利用Shafer证据理论作决策的一般方法。有了该方法以后，决策中常见的判决问题就可以得到解决。

在实践中，利用证据理论作决策时，一个首先遇到的问题就是证据从何而来？

Shafer曾经指出，在证据理论中，证据指的不是实证据，而是我们的经验和知识的一部分，是我们对该问题所作的观察和研究的结果。<sup>[1]</sup>因此对一个特定的决策问题，决策者的经验、知识以及他对该问题的观察、研究都是他用来作决策的证据。根据这些证据，他可以在决策框架上产生一个信度函数，依据本书第九章的一般方法可以得出一个他自己满意的决策结果。

但是，一个人的知识、经验等是有限的，因此决策者根据他自己的经验和知识（即证据），依据第九章的方法所作出

的决策也是有局限性的。为克服这种局限性，本书探讨了向专家咨询，利用专家的知识 and 经验来决策的可能性。

假设决策者向某专家进行咨询，那么对于该专家，他同样可以根据他自己的知识和经验（证据）在决策框架上产生一个基本可信度分配。在这里，我们作了这样一个假设：如果决策者与该专家所具有的知识和经验相同的话，那么他也可以在决策框架上产生相同的基本可信度分配。因此，在决策者向该专家进行了咨询以后，我们可以假设，决策者利用该专家所对应的经验和知识即证据在决策框架上得到了一个基本可信度分配。得到基本可信度分配以后，我们可以利用第九章的方法，首先算出该基本可信度分配所对应的信度函数，然后得出一个决策结果。该结果是决策者利用某专家的知识、经验得到的，因此具有一定的科学性。但是一个专家的经验和知识往往也带有许多片面性，因此依据他的经验和知识得出的决策结果也难免带有很大的局限性。为此可考虑向多个专家进行咨询，克服向单个专家咨询作出的决策的局限性，依靠集体的智慧作出正确的决策。Dempster合成法则为我们综合多个专家的意见提供了有力的工具。

假设决策者向多个专家进行咨询，那么他就可以根据多批证据在决策框架上得到多个基本可信度分配。依据 Dempster合成法则可求得这多个基本可信度分配的直和，从而可以利用第九章的方法得到一个决策结果。该决策结果综合了多个专家的经验和知识，反映了多个专家的意见和建议。因此这种决策具有很大的科学性，可以作为我们最后决策的结果。当然，如果我们对该结果还不满意的话，那么我们可以再向其他的专家进行咨询，我们的方法能够将这些专家的意



见加到原来的专家意见中去得出综合所有专家意见的决策结果。因此我们这种决策方法为专家集体决策提供了可能。

在第十章，我们将第九章的方法应用到决策分析领域，得到了另一种主观概率的估计方法——专家咨询法。利用这种方法可以估计出系统的各种状态发生的主观概率，从而可以依据这些主观概率作出科学的决策。

在第十一章，我们将利用 Shafer 信度理论作决策的一般方法应用到预测领域，得到了另一种预测的方法——信度预测法。信度预测法是特尔菲预测法的推广，在咨询方式的设计上以及专家经验的描述上和专家经验的处理方式上都与特尔菲法有本质的区别。

在第十二章，我们将第九章的方法应用到指标体系的确定，领导班子的选取以及医疗诊断等问题上。这些问题与前面说的主观概率的估计以及预测问题有明显的区别，因此不能直接应用第九章的方法。但是，当我们经过一些技巧性的处理以后仍可用第九章的方法进行决策。详细过程，请见该章。

与第十章，第十一章以及第十二章相类似的一些问题同样可以用第九章的方法来解决。比如：计划的制定，历史结论的确定等。社会科学包括了许多可以归结为证据处理的例子，因此，作者认为，证据理论可以作为社会科学的一种定量分析的工具。

作者认为，证据理论的另一个可适合应用的领域是人工智能和专家系统。

证据处理也是人的一项智能活动。如何将人的证据处理的能力赋予机器，也就是说如何使机器也具有证据处理的能力

力,这些研究无疑为人工智能研究提供了丰富的内容。因此,Shafer的证据理论可以看成是人工智能研究的一个方面。

作者认为,使机器具有证据处理能力的研究可以分为两个步骤:

①搞清人的证据处理的机制——证据处理的数学理论就是在这方面所作的一个尝试。

②根据这一机制逐步制造出具有证据处理能力的机器——也许与第五代计算机甚至第六代计算机有关。现在的任务是:在现有的机器上,根据现有的理论做出具有初步证据处理能力的系统。事实上,近年来在一定程度上应用Shafer的证据处理理论建造的各类专家系统就是在这一方面所取得的初步成就。

作者认为,证据处理理论不仅可以看成是人工智能的一个分支,而且也可以尝试用于人工智能研究的各个方面。本书第十三章§1、§2、§3探讨了证据理论在推理中的应用。§4探讨了证据理论在模式识别中的应用。在人类的推理中包含大量的不确定性,因此建造专家系统时必须考虑这些不确定性。Shafer信度理论是一种似然推理(Probable reasoning)的理论,这种理论从一个角度探讨了人类推理的不确定性。因此这种理论在专家系统的建造中必定可以起到重大的作用。该章§5论述了专家系统中各种不确定性的存在,综述了现有专家系统对不确定性的处理方法,探讨了应用证据理论处理专家系统中不确定性的可能性。

值得说明的是:就作者看来,Shafer证据理论不单单可以用于这些领域,在其他涉及到不确定性推理以及涉及到判决问题的领域中都可以尝试应用Shafer信度理论。

## 第二章 信度函数

### §1 识别框架

设现在有一判决问题，对于该问题我们所能认识到的所有可能的结果的集合用 $\Theta$ 表示，那么，我们所关心的任一命题都对应于 $\Theta$ 的一个子集。例如，有这样一个判决问题：今年研究生入学考试的分数线究竟是多少？对于这样一个判决问题，根据历年来分数线的变化以及所掌握的其他情况，我们首先可以确定的是分数线不会超过250到350。假设分数线只能是几百几不能是几百几十几，这样我们就得到了相对于以上判决问题的可能性集合 $\Theta = \{250, 260, \dots, 350\}$ 。在这个问题中，我们关心的无非是分数线是几或几和几，几到几，而这些都是 $\Theta$ 的子集。再如：我们要判断一件出土文物是什么朝代什么地方生产的，根据我们现有的知识和掌握的情况，可以确定这件文物或者是 $d_1$ 朝代 $p_1$ 地方生产，或者 $d_2$ 朝代 $p_2$ 地方生产， $\dots$ ，或者 $d_n$ 朝代 $p_n$ 地方生产。这样我们就得到了一个集合 $\Theta = \{<d_1, p_1>, <d_2, p_2>, \dots, <d_n, p_n>\}$ 我们所关心的命题无非是文物出产时间、地点在什么范围内，而这样一个范围正是 $\Theta$ 的一个子集。

将命题和子集对应起来可以使得我们把比较抽象的逻辑概念转化为比较直观的集论概念。事实上，任何两个命题的析取，合取和蕴含分别对应于这两个命题对应集合的并、交和包含，任何一个命题的否定对应于该命题对应集合的补。

那么，集合 $\Theta$ 到底应该怎样选择呢？

Shafer 指出： $\Theta$ 的选取依赖于我们的知识，依赖于我们的认识水平，依赖于我们所知道的和想要知道的。<sup>[1]</sup> 为了强调可能性集合 $\Theta$ 所具有的这种认识论的特性，Shafer 称其为识别框架 (frame of discernment)。而且当一个命题对应于该框架的一个子集时，称为该框架能够识别该命题。另外， $\Theta$ 的选取也应足够丰富，使我们所考虑的任何特定的命题集都可以对应于 $\Theta$ 的幂集中的某一集类。

为简单起见，本章所讨论的 $\Theta$ 都假定为有限集。关于 $\Theta$ 为无限集的情况，请参看第八章 § 3。

## §2 基本可信度分配与信度函数

有了识别框架的概念以后，我们就可以建立证据处理的数学模型了。

如何来表达人们关于一个命题的信任即相信一个命题为真的程度呢？不同的学者有不同的看法。

逻辑主义以及频率主义者认为：一个命题为真的程度是由证据完全决定的——片面强调证据的作用，忽视人的判决作用。

主观主义者认为：一个命题为真的程度完全由人决定，是人主观想象的结果——片面强调人的认识作用，忽视证据

的作用。

Shafer 认为：在一批给定的证据与一个给定的命题之间没有什么一定的客观联系能够确定一个精确的支持度；一个实在的人对于一个命题的心理描述也不是总能够用一个相当精确的实数来表示，而且也并不是总能确定这样一个数。<sup>[1]</sup>但是，对于一个命题他可以作出一种判决，在他通盘考虑了一个给定的证据组中的有时含混，有时混乱的感觉与理解之后，能够说出一个数字来表示据他本人判断出的该证据支持一个给定的命题的程度，也即他本人希望赋予该命题的那种信度。Shafer 对于人根据证据为一个命题赋予一个信度的理解可以用下列图形来表示。

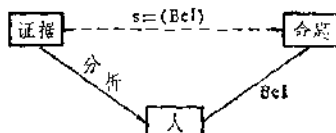


图 2.1

在证据、命题与人之间所划的实线表示人可以对证据加以分析从而得到他本人希望赋予命题的信度  $Bel$ ；在证据与命题之间所划的虚线表示一种人假想出来的证据对于命题的支持关系，是人经过证据分析后所赋予的证据对命题的支持关系，支持程度  $s = Bel$ 。所以，支持度与信度是人根据证据判断出的对命题看法的两个方面。

这种基于证据分析，确定相信一个命题为真的程度的方法，称为证据处理。

按照 Shafer 的观点, 证据处理的数学模型为:

1. 首先确立识别框架  $\Theta$ 。只有确立了框架  $\Theta$  才能使我们对于命题的研究转化为对集合的研究。

2. 根据证据建立一个信度的初始分配, 即证据处理人员对证据加以分析, 确定出证据对每一集合 (命题) 本身的支持程度 (而不去管它的任何真子集 (前因后果))。

3. 分析前因后果, 算出我们对于所有命题的信度。

从直观上看, 一批证据对一个命题提供支持的话, 那么它也应该对该命题的推论提供同样的支持。所以, 对一个命题的信度应该等于证据对它的所有前提本身提供的支持度之和。

根据证据建立的信度的初始分配我们用下面的集函数基本可信度分配 (basic probability assignment) 来表达, 对每个命题的信度用信度函数 (belief function) 来表达。

**定义 2.2.1** 设  $\Theta$  为识别框架。如果集函数  $m: 2^\Theta \rightarrow [0, 1]$  ( $2^\Theta$  为  $\Theta$  的幂集) 满足:

$$\begin{aligned} (1) \quad & m(\phi) = 0 \\ (2) \quad & \sum_{A \in 2^\Theta} m(A) = 1 \end{aligned} \tag{2.2.1}$$

则称  $m$  为框架  $\Theta$  上的基本可信度分配,  $\forall A \in 2^\Theta, m(A)$  称为  $A$  的基本可信数 (basic probability number)。

基本可信数反映了对  $A$  本身的信度大小。条件 (1) 反映了对空集 (空命题) 不产生任何信度; 条件 (2) 反映了虽然我们可以给一个命题赋任意大小的信度值, 但要求我们给所有命题赋的信度值的和等于 1, 即我们的总信度为 1。

**定义 2.2.2** 设  $\Theta$  为识别框架,  $m: 2^\Theta \rightarrow [0, 1]$  为框架  $\Theta$

上的基本可信度分配, 则称由

$$\text{Bel}(A) = \sum_{B \subset A} m(B) \quad (\forall A \subset \Theta) \quad (2.2.2)$$

所定义的函数  $\text{Bel}: 2^\Theta \rightarrow [0, 1]$  为  $\Theta$  上的信度函数。

当基本可信度分配  $m(A) = \begin{cases} 1 & A = \Theta \\ 0 & A \neq \Theta \end{cases}$  时, 信度函数的

结构是最简单的, 此时

$$\text{Bel}(A) = \begin{cases} 1 & A = \Theta \\ 0 & A \neq \Theta \end{cases}$$

该信度函数称为空信度函数(vacuous belief function)。空信度函数适合于无任何证据的情况。

**定理 2.2.1\*** 设  $\Theta$  是一个识别框架。集函数  $\text{Bel}: 2^\Theta \rightarrow [0, 1]$  是信度函数当且仅当它满足:

- (1)  $\text{Bel}(\phi) = 0$
- (2)  $\text{Bel}(\Theta) = 1$
- (3)  $\forall A_1, A_2, \dots, A_n \subset \Theta$  ( $n$  为任意自然数) (2.2.3)

$$\begin{aligned} \text{Bel}\left(\bigcup_{i=1}^n A_i\right) &\geq \sum_{i=1}^n \text{Bel}(A_i) - \sum_{1 \leq i < j \leq n} \text{Bel}(A_i \cap A_j) \\ &\quad + \dots + (-1)^{n+1} \text{Bel}\left(\bigcap_{i=1}^n A_i\right) \\ &= \sum_{I \subseteq \{1, 2, \dots, n\}} (-1)^{|I|+1} \text{Bel}\left(\bigcap_{i \in I} A_i\right) \end{aligned}$$

该定理说明, 由 (2.2.2) 式定义的函数  $\text{Bel}$  满足 (2.2.3)

---

\* 本书前七章中的定理限于篇幅一律不加证明, 有兴趣的读者可参看[1]。

式，反过来满足(2.2.3)式的函数也可以由(2.2.2)式定义，即存在函数  $m$ ，使得  $\forall A \subseteq \Theta \quad \text{Bel}(A) = \sum_{B \subseteq A} m(B)$ 。

**定理2.2.2** 设信度函数  $\text{Bel}: 2^\Theta \rightarrow [0, 1]$  是由基本可信度分配  $m: 2^\Theta \rightarrow [0, 1]$  所给出，则

$$\forall A \subseteq \Theta \quad m(A) = \sum_{B \subseteq A} (-1)^{|A-B|} \text{Bel}(B) \quad (2.2.4)$$

所以说，信度函数除了可以从基本可信度分配的角度定义以外，也可以从自身满足什么特性的角度来定义。此时(2.2.3)式可以理解成对我们的信度进行约束的规则。

为什么要给我们的信度以一些人为的约束呢？Shafer 指出：我并不要求信度一定要满足(2.2.3)式，而且也不去说如果一个人产生了不满足(2.2.3)式的信度他就是无理智的、荒谬的。不过，由于这些规则是直观的，而且对我们的理论来讲也是必须的，所以我们还是要假定我们的信度满足(2.2.3)式。<sup>[1]</sup>

说(2.2.3)式对我们的理论来说是必须的，是因为只有满足(2.2.3)式的函数才能用 Dempster 规则进行合成（见下章），换句话说，承认了我们的信度必须用 Dempster 合成法则进行合成的话，那么我们的信度就必须满足(2.2.3)式。

说只有满足(2.2.3)式的函数才能用 Dempster 合成法则进行合成并不等于说满足(2.2.3)式的信度只能用 Dempster 法则进行合成，是否还存在着与(2.2.3)式相协调的其他合成法则呢？这个问题后面还要讨论。

**定义2.2.3** 如果  $m(A) > 0$ ，则称  $A$  为信度函数  $\text{Bel}$  的



焦元(focal element)，所有焦元的并称为它的核心(core)。

**定理2.2.3** 假设  $\mathcal{C}$  是  $\Theta$  上的信度函数 Bel 的核心，那么  $\Theta$  的一个子集 B 满足  $\text{Bel}(B) = 1$  的充分必要条件为

$$\mathcal{C} \subset B$$

该定理说明，只有包含核心的集合，其信度才能达到1，否则永远小于1。

### §3 众信度函数与似真度函数

众信度函数(Commonality function)和似真度函数(plausibility function)是从另一个侧面对信度的描述。

#### 一、众信度函数

**定义2.3.1** 设函数  $Q: 2^\Theta \rightarrow [0, 1]$  由下式定义

$$\forall A \subset \Theta \quad Q(A) = \sum_{A \subset B} m(B) \quad (2.3.1)$$

则  $Q$  称为 Bel 的众信度函数。 $\forall A \subset \Theta$ ， $Q(A)$  称为 A 的众信数。

由上定义可以看出，对于一个集合(命题)  $A \subset \Theta$ ，它的众信数  $Q(A)$  反映了包含 A 的集合(被 A 所蕴含的命题)的所有基本可信数之和。Bel 是从一个结论的前提这个角度描述信度的，而  $Q$  是从一个前提的结论这个角度描述信度的。

**定理2.3.1** 设 Bel 是  $\Theta$  上的信度函数， $Q$  是它的众信度函数，那么， $\forall A \subset \Theta$

$$\text{Bel}(A) = \sum_{i=1}^n (-1)^{i+1} Q(B_i) \quad (2.3.2)$$

$$Q(A) = \sum_{B \subset A} (-1)^{|B|} \text{Bel}(\bar{B}) \quad (2.3.3)$$

该定理说明, Bel 与 Q 可以相互定义; 这也即说明, 众信度函数从另一侧面反映了信度。

众信度函数还为我们提供了另外一种研究信度函数的核心的方法。

**定理2.3.2** 设  $\mathcal{C}$  是一个信度函数的核心, 而 Q 是其对应的众信度函数, 那么  $\theta \in \Theta$  是  $\mathcal{C}$  中的元素当且仅当

$$Q(\{\theta\}) > 0$$

因为 Q 是非增的, 即若  $B \subset A$  那么  $Q(B) \geq Q(A)$  (由 Q 的定义显而易见), 所以若一个集合 A 中包含有一个不是核心中的点  $\theta$ , 那么  $Q(A)$  必等于 0。这就是说若集合 A 不完全包含于核心  $\mathcal{C}$  中的话, 它的众信数必为 0。那么若 A 完全包含于核心  $\mathcal{C}$  中的话, 情况又怎么样呢?

**例2.3.1** 设  $\Theta = \{a, b, c, d\}$ 。定义  $m: 2^\Theta \rightarrow [0, 1]$  如下

$$\forall A \subset \Theta \quad m(A) = \begin{cases} \frac{1}{3} & A = \{a, b\} \\ \frac{2}{3} & A = \{c\} \\ 0 & \text{其他} \end{cases}$$

则对应的信度函数的核心为  $\{a, b, c\}$ 。由上定理知,  $Q(\{a\}), Q(\{b\}), Q(\{c\})$  均大于 0, 但  $Q(\{a, c\}), Q(\{b, c\}), Q(\{a, b, c\})$  都等于 0。这说明, 尽管集合 A 完全包含于核心  $\mathcal{C}$ , 它的众信数也有可能等于 0。

在 (2.3.2) 式中, 令  $A = \phi$ , 得到

$$\text{Bel}(\phi) = \sum_{B \subset \phi} (-1)^{|B|} Q(B)$$

$$\text{即 } 0 = \sum_{\substack{B \subset \Theta \\ B \neq \phi}} (-1)^{|B|+1} Q(B) + (-1)^{|\Theta|+1} Q(\phi)$$

$$\text{即 } \sum_{\substack{B \subset \Theta \\ B \neq \phi}} (-1)^{|B|+1} Q(B) = 1$$

将 B 换为一般的 A 得:

$$\sum_{\substack{A \subset \Theta \\ A \neq \phi}} (-1)^{|A|+1} Q(A) = 1$$

通过该式可以很容易地计算一种特殊类型的众信度函数。

设  $q: (2^\Theta - \{\phi\}) \rightarrow [0, \infty]$  已知

而  $\forall A \subset \Theta, Q(A) = Kq(A)$  其中 K 为一未知的正常数。

将  $Q(A) = Kq(A)$  代入上式得

$$\sum_{\substack{A \subset \Theta \\ A \neq \phi}} (-1)^{|A|+1} Kq(A) = 1$$

$$\therefore K = \left( \sum_{\substack{A \subset \Theta \\ A \neq \phi}} (-1)^{|A|+1} q(A) \right)^{-1} \quad (2.3.4)$$

由此即可很容易地算出 Q。

## 二、似真度函数

关于一个命题 A 的信任单用信度函数来描述还是不够的, 因为  $\text{Bel}(A)$  不能反映出我们怀疑 A 的程度即我们相信 A 的非为真的程度。所以为了全面描述我们对 A 的信任还必须引入若干表示我们怀疑 A 的程度的量。

**定义 2.3.2** 设  $\text{Bel}: 2^\Theta \rightarrow [0, 1]$  是  $\Theta$  上的一个信度函数。定义  $\text{Dou}: 2^\Theta \rightarrow [0, 1]$  和  $\text{pl}: 2^\Theta \rightarrow [0, 1]$  如下

$$\begin{aligned} \forall A \subset \Theta \quad \text{Dou}(A) &= \text{Bel}(\bar{A}) \\ \text{pl}(A) &= 1 - \text{Bel}(\bar{A}) \end{aligned} \quad (2.3.5)$$

则称  $\text{Dou}$  为  $\text{Bel}$  的怀疑函数,  $\text{pl}$  为  $\text{Bel}$  的似真度函数。

$\forall A \subset \Theta$ ,  $\text{Dou}(A)$  称为  $A$  的怀疑度;  $\text{pl}(A)$  称为  $A$  的似真度。

$\text{Dou}(A)$  表示我们怀疑  $A$  的程度, 而  $\text{pl}(A)$  表示我们不怀疑  $A$  的程度或者说我们发现  $A$  可靠或似真的程度。由于  $\text{Dou}$  在我们今后的讨论中作用不大, 所以以后着重研究似真度函数  $\text{pl}$ 。

由 (2.3.5) 式得

$$\forall A \subset \Theta \quad \text{Bel}(A) = 1 - \text{pl}(\bar{A}) \quad (2.3.6)$$

(2.3.5) 与 (2.3.6) 式说明,  $\text{Bel}$  与  $\text{pl}$  所传递的信息完全相同, 其中的一个都可以从另一个中得到。

根据 (2.3.5) 式, 我们可以用与  $\text{Bel}$  对应的  $m$  来重新表示  $\text{pl}$

$$\begin{aligned} \forall A \subset \Theta \quad \text{pl}(A) &= 1 - \text{Bel}(\bar{A}) = \sum_{B \subset \Theta} m(B) \\ &\quad - \sum_{B \subset \bar{A}} m(B) = \sum_{B \cap A \neq \emptyset} m(B) \end{aligned} \quad (2.3.7)$$

若  $A \cap B \neq \emptyset$ , 则称  $A$  与  $B$  相容。(2.3.7) 式说明,  $\text{pl}(A)$  包含了所有与  $A$  相容的那些集合 (命题) 的基本可信数。

对比 (2.2.2) 式与 (2.3.7) 式, 我们立即可以知道

$$\forall A \subset \Theta \quad \text{Bel}(A) \leq \text{pl}(A) \quad (2.3.8)$$

意即:  $\text{pl}(A)$  是比  $\text{Bel}(A)$  更宽松的一种估计, 或  $\text{Bel}(A)$  是比  $\text{pl}(A)$  更保守的一种估计。

对比 (2.3.1) 式与 (2.3.7) 式, 我们可以知道

$$\forall \theta \in \Theta \quad \text{pl}(\{\theta\}) = Q(\{\theta\}) \quad (2.3.9)$$

该式说明，在单点集上似真度与众信度是相同的。

另外，由下面定理我们也可以知道，似真度函数和众信度函数也可以互相定义。

**定理2.3.3** 设  $\text{Bel}$  是  $\Theta$  上的信度函数， $\text{pl}$  和  $Q$  分别是  $\text{Bel}$  所对应的似真度函数和众信度函数。那么

$$\forall A \subset \Theta \quad A \neq \emptyset$$

$$\text{pl}(A) = \sum_{\substack{B \subset A \\ B \neq \emptyset}} (-1)^{|B|+1} Q(B) \quad (2.3.10)$$

$$Q(A) = \sum_{B \subset A} (-1)^{|B|+1} \text{pl}(B) \quad (2.3.11)$$

由 (2.2.2)，(2.2.4)，(2.3.2)，(2.3.3) 以及 (2.3.5)，(2.3.6)，(2.3.10)，(2.3.11) 式可知： $m, \text{Bel}, \text{pl}, Q$  这四个函数是一一对应的，只要知道其中之一，其余的都可以计算出来。

#### §4 $m, \text{Bel}, \text{pl}, Q$ 的几何意义

本节讨论这几个函数的几何意义。

如果我们将  $\Theta$  的元素看成点，那么我们就可以将我们的信度看成是一种半流动的“信质” (Semi-mobile “probability masses”)，这些信质可以从一点移动到另一点，但是要局限于  $\Theta$  的各个子集中。

$m(A)$  反映了对  $A$  本身 (而不管前因后果) 的信度大小，即反映了局限于  $A$  而不局限于  $A$  的任何真子集的总信质。换句话说， $m(A)$  是局限于  $A$  中可以自由移动到  $A$  的每一点的信质。

$Bel(A) = \sum_{B \subset A} m(B)$  是分配到  $A$  上的总信度，它综合了

$A$  的所有前提本身的信度，所以， $Bel(A)$  是局限于  $A$ ，可以在  $A$  中自由移动但不一定能达到每一点的总信质。

$Q(A) = \sum_{A \subset B} m(B)$  是  $A$  的所有结论本身的信度之和，所

以  $Q(A)$  是可以自由移动到  $A$  的每一点但不一定局限于  $A$  中的总信质。

$pl(A) = \sum_{B \cap A \neq \emptyset} m(B)$  是所有与  $A$  相容的命题本身的信度之

和，所以  $pl(A)$  是可以自由移动到  $A$  但不一定局限于  $A$  也不一定可以自由移动到  $A$  的每一点上的总信质。

## §5 贝叶斯信度函数

定理 2.2.1 的第(3)条说明信度函数不一定满足可加性而满足一种所谓的半可加性。显然，满足可加必满足半可加。换句话说，满足可加的函数是信度函数的特例。

**定义 2.5.1** 如果  $\Theta$  是一个框架， $Bel: 2^\Theta \rightarrow [0, 1]$  满足

$$(1) Bel(\emptyset) = 0$$

$$(2) Bel(\Theta) = 1$$

$$(3) \text{ 若 } A, B \subset \Theta \text{ 且 } A \cap B = \emptyset \text{ 那么}$$

$$Bel(A \cup B) = Bel(A) + Bel(B)$$

则  $Bel$  称为贝叶斯信度函数。

**定理 2.5.1** 贝叶斯信度函数是信度函数。

**定理 2.5.2** 设  $Bel: 2^\Theta \rightarrow [0, 1]$  是信度函数， $pl$  是其对应的似真度函数，那么下列结论彼此等价：

- (1) Bel 是贝叶斯信度函数;
- (2) Bel 的所有焦元都是单点;
- (3) 包含一个以上元素的任何  $\Theta$  的子集都具有 0 众信数;
- (4)  $\text{Bel} = \text{pl}$ ;
- (5)  $\forall A \subset \Theta, \text{Bel}(A) + \text{Bel}(\bar{A}) = 1$

该定理的(5)说明, 我们的信度要么局限于  $A$ , 要么局限于  $\bar{A}$ 。换句话说, 对于一个命题, 我们要么相信该命题本身, 要么相信这个命题的非。而这个特点正是贝叶斯理论的核心。这就是为什么我们将此种特殊类型的信度函数称作贝叶斯信度函数的原因。

只要一个信度函数的焦元是单点的话, 那么所对应的信质就不能自由移动, 而只能呆在某一单点上。所以定理 2.5.2 的(2)说明, 对贝叶斯信度函数而言, 信质无任何自由移动性可言。

**定理 2.5.3** 一个信度函数  $\text{Bel}: 2^\Theta \rightarrow [0, 1]$  是贝叶斯信度函数当且仅当存在一个唯一的函数  $p: \Theta \rightarrow [0, 1]$  使得

$$\forall A \subset \Theta$$

$$\begin{aligned} \text{满足} \quad & \sum_{\theta \in \Theta} p(\theta) = 1 \\ \text{以及} \quad & \text{Bel}(A) = \sum_{\theta \in A} p(\theta) \end{aligned}$$

该定理说明, 贝叶斯信度函数与一般的信度函数相比是简单明了, 极易描述的。因为定理 2.5.2 的(2)和定理 2.5.3 说明, 贝叶斯信度函数实质上是一个点函数而不是集函数。

也正因为它简单, 所以使得我们对证据的描述是极其粗劣的甚至是不适合的。这一点在第五章中会看得更清楚。

### 第三章 Dempster 合成法则

Dempster 合成法则是一个反映证据的联合作用的一个法则。给定几个同一识别框架上基于不同证据的信度函数, 如果这几批证据不是完全冲突的, 那么我们就可以利用 Dempster 合成法则计算出一个信度函数, 而这个信度函数就可以作为在那几批证据的联合作用下产生的信度函数。该信度函数称为原来那几个信度函数的直和。

早在1764年, Lambert 就研究了在识别框架中只包含有两个元素的特殊情况下, Dempster 合成法则的陈述和使用情况。但是200年以来, Dempster 合成法则的一般情况却从来无人问津。直到1967年, Dempster 才针对统计推理首先研究并给出了这种一般的合成法则。<sup>[16]</sup>1976年, Shafer 对此又进行了深入的研究和分析, 并给出了针对一般情况的更加一般的合成法则。

#### §1 两个信度函数的合成

设  $Bel_1$  和  $Bel_2$  是同一识别框架  $\Theta$  上的两个信度函数,  $m_1$  和  $m_2$  分别是其对应的基本可信度分配, 焦点分别为  $A_1, \dots$ ,



$A_k$ 和 $B_1, \dots, B_l$ , 分别用图3.1和图3.2来表示。

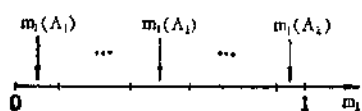


图 3.1

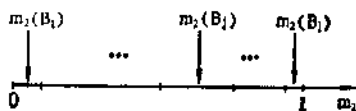


图 3.2

在图3.1和图3.2中,  $[0, 1]$ 中的某一段表示由各自的基本可信度分配决定的某一焦元上的信质, 并不表示整个的识别框架。

如图3.3, 我们将图3.1和图3.2综合起来考虑可得到一

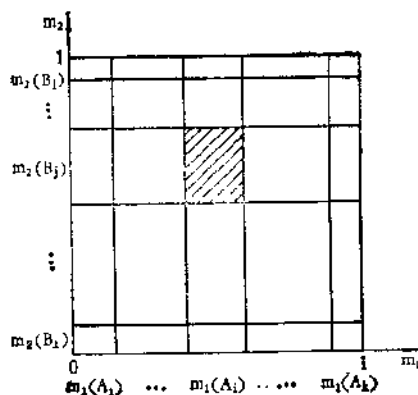


图 3.3

系列的矩形。将整个大矩形看作我们的总的信质, 那么, 一个一个的竖条就表示 $m_1$ 分配到它的焦元 $A_1, \dots, A_k$ 上的信质。同理, 一个一个的横条就表示 $m_2$ 分配到它的焦元 $B_1, \dots, B_l$ 上的信质。如图, 一个横条和一个竖条的交是一个小矩形,

该矩形具有测度  $m_1(A_i)m_2(B_j)$ 。因为它是同时分配到  $A_i$  和  $B_j$  上的,所以我们可以说,  $Bel_1$  和  $Bel_2$  的联合作用就是将  $m_1(A_i)m_2(B_j)$  确切地分配到  $A_i \cap B_j$  上。

给定  $A \subset \Theta$ , 若有  $A_i \cap B_j = A$ , 那么  $m_1(A_i)m_2(B_j)$  就是确切地分配到  $A$  上的一部分信质, 所以确切地分配到  $A$  上的总信质为

$$\sum_{A_i \cap B_j = A} m_1(A_i)m_2(B_j)$$

但是当  $A = \phi$  时, 按这种理解, 我们将有一部分信质  $\sum_{A_i \cap B_j = \phi} m_1(A_i)m_2(B_j)$  分配到空集上, 这显然是不合理的。怎么办呢? 唯一的办法就是丢弃这部分信质, 但丢弃这部分信质以后, 我们的总信质就要小于 1, 为此还需在每一信质上乘一系数

$$\left(1 - \sum_{A_i \cap B_j = \phi} m_1(A_i)m_2(B_j)\right)^{-1} \quad (3.1.1)$$

以满足总信质为 1 的要求。

**定理 3.1.1** 设  $Bel_1$  和  $Bel_2$  是同一识别框架  $\Theta$  上的两个信度函数,  $m_1$  和  $m_2$  分别是其对应的基本可信度分配, 焦元分别为  $A_1, \dots, A_n$  和  $B_1, \dots, B_l$ , 设

$$\sum_{A_i \cap B_j \neq \phi} m_1(A_i)m_2(B_j) < 1 \quad (3.1.2)$$

那么 由下式定义的函数  $m: 2^\Theta \rightarrow [0, 1]$  是基本可信度分配

$$m(A) = \begin{cases} 0 & A = \phi \\ \frac{\sum_{A_i \cap B_j = A} m_1(A_i)m_2(B_j)}{1 - \sum_{A_i \cap B_j = \phi} m_1(A_i)m_2(B_j)} & A \neq \phi \end{cases} \quad (3.1.3)$$

由 $m$ 所给定的信度函数的核心等于 $Bel_1$ 和 $Bel_2$ 的核心的交。由 $m$ 给定的信度函数称为 $Bel_1$ 和 $Bel_2$ 的直和, 记为 $Bel_1 \oplus Bel_2$ 。如果(3.1.2)式不成立, 那么就说 $Bel_1$ 与 $Bel_2$ 的直和 $Bel_1 \oplus Bel_2$ 不存在。

**定理3.1.2** 设 $Bel_1$ 和 $Bel_2$ 是同一识别框架 $\Theta$ 上的两个信度函数,  $Q_1$ 和 $Q_2$ 分别表示它们的众信度函数, 那么下列条件都是等价的

- (1)  $Bel_1 \oplus Bel_2$ 不存在;
- (2)  $Bel_1$ 和 $Bel_2$ 的核心不相交;
- (3)  $\exists A \subset \Theta$ 使 $Bel_1(A) = 1$ 且 $Bel_2(\bar{A}) = 1$ ;
- (4)  $\forall A \subset \Theta, A \neq \phi, Q_1(A)Q_2(A) = 0$ 。

该定理的(2)表明当 $Bel_1$ 和 $Bel_2$ 的核心不相交时,  $Bel_1$ 与 $Bel_2$ 不能合成,  $Bel_1$ 和 $Bel_2$ 核心不相交也即它们所对应的证据支持完全不同的命题即是两批完全不同的证据, 而这样的两批证据显然是合不到一块的。定理的(3)表明当这两批证据一个完全支持命题, 一个完全支持该命题的非时, 这两批证据也是不能合成的, 这一点也符合我们的直观理解。定理的(4)的意义从下节的定理中一看即明。

定理3.1.2从另一个侧面说明Dempster合成法则可以反映两批证据的联合作用。

易见, 当 $Bel_1 \oplus Bel_2$ 存在时,  $Bel_2 \oplus Bel_1$ 也存在, 且

$$Bel_1 \oplus Bel_2 = Bel_2 \oplus Bel_1$$

**定理3.1.3** 设 $Bel_1$ 和 $Bel_2$ 是框架 $\Theta$ 上的信度函数, 那么

- (1) 如果 $Bel_1$ 是空信度函数, 那么 $Bel_1$ 与 $Bel_2$ 是可合成的, 并且 $Bel_1 \oplus Bel_2 = Bel_2$ ;
- (2) 如果 $Bel_1$ 是贝叶斯信度函数而且 $Bel_1$ 和 $Bel_2$ 是可合

成的, 那么  $\text{Bel}_1 \oplus \text{Bel}_2$  是贝叶斯信度函数。

## §2 多个信度函数的合成

设  $\text{Bel}_1, \dots, \text{Bel}_n$  是同一识别框架  $\Theta$  上的几个信度函数, 若  $\text{Bel}_1$  与  $\text{Bel}_2$  是可合成的, 那么  $\text{Bel}_1 \oplus \text{Bel}_2$  也是一个信度函数; 若该信度函数与  $\text{Bel}_3$  也是可合成的, 那么  $(\text{Bel}_1 \oplus \text{Bel}_2) \oplus \text{Bel}_3$  仍然是一个信度函数, 而且由下节定理 3.3.2 可知,  $(\text{Bel}_1 \oplus \text{Bel}_2) \oplus \text{Bel}_3 = \text{Bel}_1 \oplus (\text{Bel}_2 \oplus \text{Bel}_3)$ ; 由此可定义  $\text{Bel}_1 \oplus \text{Bel}_2 \oplus \text{Bel}_3$ 。同样可依次类推定义  $\text{Bel}_1 \oplus \dots \oplus \text{Bel}_n$ 。

若  $\text{Bel}_1 \oplus \dots \oplus \text{Bel}_n$  存在, 那么在构造它的过程中 计算信度函数的直和的顺序如何就是无关紧要的。与两个信度函数的直和可以表示这两个信度函数对应证据的联合作用一样, 多个信度函数的直和也可以表示这多个信度函数对应证据的联合作用。

$\text{Bel}_1 \oplus \dots \oplus \text{Bel}_n$  可以一个一个地依次计算, 也可以直接根据  $\text{Bel}_1, \dots, \text{Bel}_n$  所对应的  $m_1, \dots, m_n$  计算出它所对应的  $m$ 。

**定理 3.2.1** 设  $\text{Bel}_1, \dots, \text{Bel}_n$  是同一识别框架  $\Theta$  上的信度函数,  $m_1, \dots, m_n$  是对应的基本可信度分配, 如果  $\text{Bel}_1 \oplus \dots \oplus \text{Bel}_n$  存在且基本可信度分配为  $m$ , 则

$$\begin{aligned} \forall A \subseteq \Theta \quad A \neq \emptyset \\ m(A) = K \sum_{\substack{A_1, \dots, A_n \subseteq \Theta \\ A_1 \cap \dots \cap A_n = A}} m_1(A_1) \cdots m_n(A_n) \end{aligned} \quad (3.2.1)$$

$$\text{其中 } K = \left( \sum_{\substack{A_1, \dots, A_n \subset \Theta \\ A_1 \cap \dots \cap A_n \neq \emptyset}} m_1(A_1) \cdots m_n(A_n) \right)^{-1} \quad (3.2.2)$$

此即多个信度函数合成的Dempster法则。

该定理的前提条件是 $\text{Bel}_1 \oplus \cdots \oplus \text{Bel}_n$ 存在,那么在什么情况下 $\text{Bel}_1 \oplus \cdots \oplus \text{Bel}_n$ 存在呢?

**定理3.2.2** 设 $\text{Bel}_1, \dots, \text{Bel}_n$ 是同一识别框架 $\Theta$ 上的信度函数, $\mathcal{C}_1, \dots, \mathcal{C}_n$ 分别是其对应的核心, $Q_1, \dots, Q_n$ 分别是其对应的众信度函数。那么下列条件是等价的

- (1)  $\mathcal{C}_1 \cap \cdots \cap \mathcal{C}_n = \emptyset$ ;
- (2)  $\forall A \subset \Theta, A \neq \emptyset, Q_1(A) \cdots Q_n(A) = 0$ ;
- (3)  $\exists A_1, \dots, A_n \subset \Theta$  满足  $A_1 \cap \cdots \cap A_n = \emptyset$  且  
 $\text{Bel}_i(A_i) = 1 \quad (i = 1, \dots, n)$ ;

(4)  $\exists A, B \subset \Theta$  满足  $A \cap B = \emptyset$  且有  $E_1, \dots, E_r$  和  $G_1, \dots, G_s$ , 使得

$$\begin{aligned} A &= E_1 \cap \cdots \cap E_r \\ B &= G_1 \cap \cdots \cap G_s \end{aligned}$$

而且对每一个  $E_i$  或  $G_j (i = 1, \dots, r; j = 1, \dots, s)$  至少存在一个  $\text{Bel}_i$  满足  $\text{Bel}_i(E_i) = 1$  或  $\text{Bel}_i(G_j) = 1$ ;

(5) 依靠任何顺序计算  $\text{Bel}_1 \oplus \cdots \oplus \text{Bel}_n$  的尝试都是失败的。

对于一个识别框架 $\Theta$ 上的 $\text{Bel}_1, \dots, \text{Bel}_n$ ,若不满足定理3.2.2中的条件,就称为是可合成的。此时, $\text{Bel}_1 \oplus \cdots \oplus \text{Bel}_n$ 是存在的并可称为 $\text{Bel}_1, \dots, \text{Bel}_n$ 的直和。

与定理3.1.2相类似,同样可讨论定理3.2.2中各条的意义。

### §3 Dempster合成法则的 众信度函数形式

Dempster合成法则用众信度函数来表示具有更清晰和更整齐的形式。

**定理3.3.1** 设 $Bel_1$ 和 $Bel_2$ 是识别框架 $\Theta$ 上的两个信度函数,  $Bel_1 \oplus Bel_2$ 存在。  $Q_1$ ,  $Q_2$ 和 $Q$ 分别表示 $Bel_1$ ,  $Bel_2$ 和 $Bel_1 \oplus Bel_2$ 的众信度函数, 那么

$$\forall A \subset \Theta \quad Q(A) = K Q_1(A) Q_2(A)$$

其中,  $K$ 不依赖于 $A$

根据 (2.3.4) 式可得

$$K = \left( \sum_{\substack{B \subset \Theta \\ B \neq \emptyset}} (-1)^{|B|+1} Q_1(B) Q_2(B) \right)^{-1} \quad (3.3.1)$$

而且从本定理的证明中可知  $K$ 与本章 (3.1.1)式归一化常数相同。

**定理3.3.2** 设 $Bel_1, \dots, Bel_n$ 是同一识别框架 $\Theta$ 上的信度函数,  $Bel_1 \oplus \dots \oplus Bel_n$ 存在,  $Q_1, \dots, Q_n$ 和 $Q$ 分别是 $Bel_1, \dots, Bel_n$ 和 $Bel_1 \oplus \dots \oplus Bel_n$ 所对应的众信度函数, 那么

$$\forall A \subset \Theta \quad A \neq \emptyset$$

$$Q(A) = K Q_1(A) \dots Q_n(A) \quad (3.3.2)$$

其中 $K$ 不依赖于 $A$ 的变化

同理, 由 (2.3.4) 式可得

$$K = \left( \sum_{\substack{B \subset \Theta \\ B \neq \emptyset}} (-1)^{|B|+1} Q_1(B) \dots Q_n(B) \right)^{-1} \quad (3.3.3)$$

并且该常数与本章 (3.2.2) 式的归一化常数K也是相同的。

#### §4 冲突权重

两批证据在同一识别框架上的作用, 只要不是完全一致的, 就要表现出若干冲突的特性。

设  $Bel_1$  和  $Bel_2$  是识别框架  $\Theta$  上的两个信度函数。在用 Dempster 合成法则求它们的直和时, 曾经说过当  $A_1 \cap B_1 = \phi$  时丢弃  $m_1(A_1)m_2(B_1)$ 。在  $Bel_1$  与  $Bel_2$  合成的整个过程中所丢弃的总的信质为

$$k = \sum_{A_1 \cap B_1 = \phi} m_1(A_1)m_2(B_1)$$

当  $A_1 \cap B_1 = \phi$  时,  $Bel_1$  与  $Bel_2$  分别给两个不相容的命题赋予了信度, 这也就是说  $Bel_1$  与  $Bel_2$  所对应的两批证据在这个问题上发生了冲突。 $k$  值越大, 说明这两批证据间的冲突也越大; 反之,  $k$  值越小, 说明这两批证据间的冲突也越小。所以,  $k$  值可以反映这两批证据间的冲突。

(3.1.1) 式的归一化常数  $K = 1/(1 - k)$  是  $k$  的递增函数。所以, 归一化常数  $K$  也可用来表示  $Bel_1$  与  $Bel_2$  间的冲突程度。但是最适合于表示两者间冲突程度的量是

$$\log K = \log \frac{1}{1 - k} = -\log(1 - k)$$

记为  $Con(Bel_1, Bel_2)$ 。因为  $K$  总是大于等于 1, 所以  $Con(Bel_1, Bel_2)$  可以取任何大于等于 0 的值。如果  $Bel_1$  与  $Bel_2$  一点也不冲突, 那么  $k = 0$  而  $Con(Bel_1, Bel_2) = 0$ ; 如果  $Bel_1$  与  $Bel_2$  直接冲突以至于  $Bel_1 \oplus Bel_2$  不存在的话, 那么  $k = 1$  而

$\text{Con}(\text{Bel}_1, \text{Bel}_2) = \infty$ ,  $\text{Con}(\text{Bel}_1, \text{Bel}_2)$  称为  $\text{Bel}_1$  与  $\text{Bel}_2$  间的冲突权重 (Weight of conflict)。

同样, 当  $\text{Bel}_1, \dots, \text{Bel}_n$  是同一识别框架  $\Theta$  上的  $n$  个信度函数时, 可定义常数

$$\log K = \log \frac{1}{1-k} = -\log(1-k).$$

其中 
$$k = \sum_{\substack{A_1, \dots, A_n \in \Theta \\ A_i \cap A_j = \emptyset \\ A_i \cap A_k = \emptyset}} m_1(A_1) \cdots m_n(A_n)$$

为  $\text{Bel}_1, \dots, \text{Bel}_n$  间的冲突权重。记为  $\text{Con}(\text{Bel}_1, \dots, \text{Bel}_n)$ 。冲突权重也满足可加性。

**定理 3.4.1** 设  $\text{Bel}_1, \dots, \text{Bel}_n$  是识别框架  $\Theta$  上的信度函数,  $\text{Bel}_1 \oplus \dots \oplus \text{Bel}_n$  存在, 那么

$$\begin{aligned} \text{Con}(\text{Bel}_1, \dots, \text{Bel}_{n+1}) &= \text{Con}(\text{Bel}_1, \dots, \text{Bel}_n) \\ &\quad + \text{Con}(\text{Bel}_1 \oplus \dots \oplus \text{Bel}_n, \\ &\quad \text{Bel}_{n+1}) \end{aligned}$$

## §5 条件信度函数

如前所述, Shafer 证据理论对于新证据的加入采用合成法则而不采用条件法则。但是, 由于这种理论以贝叶斯理论为其特例, 所以说, 当新证据的加入具有某些特殊性时, 也可以导出一种所谓的条件法则。

设新证据在框架  $\Theta$  上的作用是确切地支持  $\Theta$  的一个子集  $B$ , 那么  $\text{Bel}_2$  将把信度 1 赋予  $B$  所对应的命题以及由  $B$  所蕴含的命题, 即



$$\text{Bel}_2(A) = \begin{cases} 1 & B \subset A \\ 0 & B \not\subset A \end{cases} \quad (3.5.1)$$

当 $\text{Bel}_2$ 取上式形式时,设 $\text{Bel}_1$ 是同一识别框架 $\Theta$ 上的一个信度函数,那么 $\text{Bel}_1 \oplus \text{Bel}_2$ 可以看成是在 $B$ 成立的条件下,集合的条件信度函数。

由Dempster合成法则可以得到条件信度的简单计算式。

**定理3.5.1** 设 $\text{Bel}_2$ 是由(3.5.1)式所定义的框架 $\Theta$ 上的信度函数, $\text{Bel}_1$ 是框架 $\Theta$ 上的另一信度函数,那么 $\text{Bel}_1$ 与 $\text{Bel}_2$ 可合成的充分必要条件是 $\text{Bel}_1(\bar{B}) < 1$ 。如果 $\text{Bel}_1$ 与 $\text{Bel}_2$ 是可合成的,令 $\text{Bel}_1(\cdot|B)$ 表示 $\text{Bel}_1 \oplus \text{Bel}_2$ ,  $\text{pl}_1$ 和 $\text{pl}_1(\cdot|B)$ 分别表示 $\text{Bel}_1$ 和 $\text{Bel}_1 \oplus \text{Bel}_2$ 的似真度函数。那么

$$\forall A \subset \Theta$$

$$\begin{aligned} \text{Bel}_1(A|B) &= \frac{\text{Bel}_1(A \cup \bar{B}) - \text{Bel}_1(\bar{B})}{1 - \text{Bel}_1(\bar{B})} \\ \text{pl}_1(A|B) &= \frac{\text{pl}_1(A \cap B)}{\text{pl}_1(B)} \end{aligned} \quad (3.5.2)$$

(3.5.2)式称为Dempster条件法则。

在第十三章的§1,我们还要讨论Dempster条件法则,并得到了Dempster条件法则的众信度形式。在那里,我们还利用Dempster条件法则讨论了某类近似推理——软推理的描述问题。

## 第四章 框架的转化

一般说来,对任何一个证据处理人员来讲,他所能想到的各种有用的观念、判断是非常多的,某一框架只能包含其中很少的一部分,所以用单个识别框架来处理问题是很不够的。在这种情况下证据处理人员为了进行某种特殊的似真推理(Probable reasoning),就有必要求助于不同的观念,侧重于不同的特性相应地改换他所采用的识别框架。粗化与细分就是为适应这个要求采用的两种变换方法。值得说明的是框架的转化不是随意的。尽管转化前后两者要强调不同的特性(distinction)、不同的侧面,而且在其关注的方向上可以具有不同程度的判决,但是两者决不能使用相互矛盾的,互不相容的概念。细分(Refinement)与粗化(Coarsening)就是两种相容的变换,而收缩(abridgment)与扩张(enlargement)就不是相容变换。

### §1 粗化与细分

设 $\mathcal{Q}$ 为一个识别框架,它的元素是由另外一个识别框架 $\mathcal{Q}^*$ 中的元素剖分而得到的。对任意 $\theta \in \mathcal{Q}$ ,将所有由 $\theta$ 通过剖分

得到的那些可能事件或结果记为  $\omega(\{\theta\})$ , 则  $\omega(\{\theta\}) \subset \Omega$ 。  
我们假定所有的  $\omega(\{\theta\})$  ( $\theta \in \Theta$ ) 组成了  $\Omega$  的一个划分。即

- (1)  $\forall \theta \in \Theta, \omega(\{\theta\}) \neq \phi$
- (2) 如果  $\theta \neq \theta'$  则  $\omega(\{\theta\}) \cap \omega(\{\theta'\}) = \phi$
- (3)  $\bigcup_{\theta \in \Theta} \omega(\{\theta\}) = \Omega$

则  $\forall A \subset \Theta$ , 令  $\omega(A) = \bigcup_{\theta \in A} \omega(\{\theta\})$ , 那么  $\omega(A)$  包含了所有通过剖分  $A$  的元素得到的可能结果。我们称这样定义的映射  $\omega: 2^\Theta \rightarrow 2^\Omega$  为  $\Theta$  到  $\Omega$  的一个细分。如果在  $\Theta$  和  $\Omega$  之间存在这样一个细分的话, 那么  $\Omega$  就称为  $\Theta$  的精细, 而  $\Theta$  称为  $\Omega$  的粗化。

框架  $\Omega$  可以识别的命题包括了所有  $\Omega$  的粗化框架  $\Theta$  可以识别的命题。事实上, 由子集  $A \subset \Theta$  表示的命题也可以由子集  $\omega(A) \subset \Omega$  来表示, 所以当  $2^\Theta$  和  $2^\Omega$  看成是命题的集合而不是子集的集合后,  $2^\Theta$  将是  $2^\Omega$  的一个子集。

用数学语言可将细分的概念陈述如下。

**定义 4.1.1** 设  $\Theta$  和  $\Omega$  是两个有限集合, 定义映射  $\omega: 2^\Theta \rightarrow 2^\Omega$  满足

(1) 对所有的  $\theta \in \Theta$ , 所有的  $\omega(\{\theta\})$  构成了  $\Omega$  的一个划分;

$$(2) \forall A \subset \Theta, \omega(A) = \bigcup_{\theta \in A} \omega(\{\theta\});$$

则称映射  $\omega$  为从  $\Theta$  到  $\Omega$  的一个细分。

**例 4.1.1** 设  $\Theta = \{\theta_1, \theta_2, \theta_3, \theta_4, \theta_5, \theta_6\}$ 。如下页图 4.1 所示, 映射  $\omega: 2^\Theta \rightarrow 2^\Omega$  将  $\Omega$  划分成 6 个部分,  $\omega$  即为从  $\Theta$  到  $\Omega$  的一个细分。

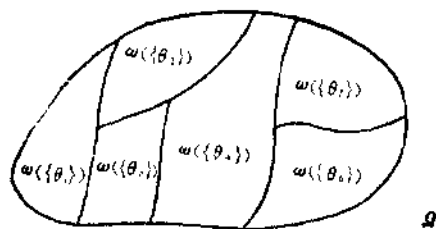


图 4.1

下面的定理反映了细分的一些性质。

**定理 4.1.1** 设  $\omega: 2^\Theta \rightarrow 2^\Omega$  是一个从  $\Theta$  到  $\Omega$  的细分, 则

- (1)  $\omega$  是一对一的;
- (2)  $\omega(\emptyset) = \emptyset$ ;
- (3)  $\omega(\Theta) = \Omega$ ;
- (4)  $\forall A, B \subset \Theta, \omega(A \cup B) = \omega(A) \cup \omega(B)$ ;
- (5)  $\forall A \subset \Theta, \omega(\bar{A}) = \overline{\omega(A)}$ ;
- (6)  $\forall A, B \subset \Theta, \omega(A \cap B) = \omega(A) \cap \omega(B)$ ;
- (7) 设  $A, B \subset \Theta$ , 那么  $\omega(A) \subset \omega(B)$  的充分必要条件是  $A \subset B$ ;

是  $A \subset B$ ;

- (8) 设  $A, B \subset \Theta$ , 那么  $\omega(A) \cap \omega(B) = \emptyset$  的充分必要条件是  $A \cap B = \emptyset$ 。

对于粗化与细分还有下列内、外收缩的概念。

**定义 4.1.2** 设映射  $\omega: 2^\Theta \rightarrow 2^\Omega$  是从  $\Theta$  到  $\Omega$  的一个细分。

定义映射  $\bar{\omega}: 2^\Omega \rightarrow 2^\Theta$  满足

$$\forall A \subset \Omega \quad \bar{\omega}(A) = \{\theta \mid \theta \in \Theta, \omega(\{\theta\}) \subset A\}$$

映射  $\bar{\omega}: 2^\Omega \rightarrow 2^\Theta$  满足

$$\forall A \subset \Omega \quad \bar{\omega}(A) = \{\theta \mid \theta \in \Theta, \omega(\{\theta\}) \cap A \neq \emptyset\}$$

则称  $\bar{\omega}$  和  $\bar{\omega}$  分别为细分  $\omega$  的内、外收缩。

所以说，一个从 $\Theta$ 到 $\Omega$ 的细分 $\omega$ 的内收缩是一个从 $\Omega$ 到 $\Theta$ 的映射，该映射满足它的像点中的点经 $\omega$ 映射后包含在 $A$ 中，这也就是为什么称 $\theta$ 为内收缩的原因。而外收缩是一个从 $\Omega$ 到 $\Theta$ 的映射，该映射满足它的像点中的点经 $\omega$ 映射后与 $A$ 相容，即映射过去再映射回来后能将把 $A$ 包起来，所以称 $\bar{\theta}$ 为细分 $\omega$ 的外收缩。细分 $\omega$ 从 $\Theta$ 映到 $\Omega$ ，而内、外收缩都是从 $\Omega$ 映到 $\Theta$ 。与细分 $\omega$ 相同，这两种映射在今后的讨论中也是相当有用的。

例4.1.2 设对于例4.1.1，有一集合 $A \subset \Omega$ 。如图4.2中的阴影部分所示，则

$$\theta(A) = \{\theta_3\}$$

$$\bar{\theta}(A) = \{\theta_1, \theta_2, \theta_3, \theta_4\}$$

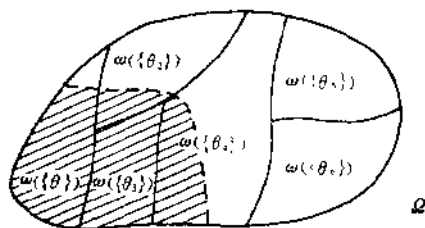


图 4.2

下面的定理使得我们可以给出 $\theta(A)$ 和 $\bar{\theta}(A)$ 的一个更加直观的解释。

**定理4.1.2** 设 $\omega: 2^\Theta \rightarrow 2^\Omega$ 是一个细分， $A \subset \Omega$ ， $B \subset \Theta$ ， $\theta$ 和 $\bar{\theta}$ 是 $\omega$ 的内、外收缩，那么 $\omega(B) \subset A$ 的充分必要条件为 $B \subset \theta(A)$ ； $A \subset \omega(B)$ 的充分必要条件为 $\bar{\theta}(A) \subset B$ 。

因为 $B$ 和 $\omega(B)$ 对应同一命题，所以 $\omega(B) \subset A$ 意味着 $B$ 蕴含 $A$ 。由此，根据上定理， $B$ 蕴含 $A$ 的充分必要条件为 $B \subset \theta(A)$ 。

换句话说,若一个集合不包含于 $\underline{\theta}(A)$ ,则该集合不能蕴含 $A$ 。另外,由 $\underline{\theta}(A)$ 的定义, $\underline{\theta}(A)$ 蕴含 $A$ ,所以说 $\underline{\theta}(A)$ 是蕴含 $A$ 的 $\Theta$ 的最大子集。同样,由定理4.1.2, $A$ 蕴含 $B$ 的充分必要条件为 $B \supset \bar{\theta}(A)$ 。换句话说,若一个集合不能包含 $\bar{\theta}(A)$ ,则该集合就不能被 $A$ 所蕴含。又由 $\bar{\theta}(A)$ 的定义, $A$ 蕴含 $\bar{\theta}(A)$ ,所以说 $\bar{\theta}(A)$ 是被 $A$ 蕴含的 $\Theta$ 的最小子集。

内、外收缩还满足定理4.1.3列出的性质。

**定理4.1.3** 设 $\omega:2^\Theta \rightarrow 2^\Theta$ 是一个细分, $\underline{\theta}$ 和 $\bar{\theta}$ 分别表示它的内、外收缩。则

- (1)  $\forall A \subset \Omega, \underline{\theta}(A) \subset \bar{\theta}(A)$ ;
- (2)  $\forall A \subset \Theta, \underline{\theta}(\omega(A)) = \bar{\theta}(\omega(A)) = A$ ;
- (3)  $\forall A \subset \Omega, \omega(\underline{\theta}(A)) \subset A$ ;
- (4)  $\forall A \subset \Omega, A \subset \omega(\bar{\theta}(A))$ ;
- (5)  $\underline{\theta}(\phi) = \bar{\theta}(\phi) = \phi$ ;
- (6)  $\underline{\theta}(\Omega) = \bar{\theta}(\Omega) = \Theta$ ;
- (7) 若 $A \subset B$ ,那么 $\underline{\theta}(A) \subset \underline{\theta}(B)$ ,  $\bar{\theta}(A) \subset \bar{\theta}(B)$ ;
- (8)  $\forall A, B \subset \Omega, \underline{\theta}(A \cap B) = \underline{\theta}(A) \cap \underline{\theta}(B)$ ;
- (9)  $\forall A, B \subset \Omega, \bar{\theta}(A \cup B) = \bar{\theta}(A) \cup \bar{\theta}(B)$ ;
- (10)  $\forall A \subset \Theta, \overline{\underline{\theta}(A)} = \bar{\theta}(A)$ 。

内、外收缩的这些性质与粗糙集中集合的上、下近似具有很大的相似性,参见第八章 §3。

## §2 相容框架族

对于框架细分来讲,是否存在一个极限化的精细框架(Ultimate Refinement),这是Shafer理论与贝叶斯理论的

根本区别。贝叶斯理论假定这样一个框架存在，即存在一个不能再细分的框架，而Shafer明确指出，我们必须舍弃这样一个假设，即在我们的理论中，任何一个被细分了的框架都可以继续进行细分。<sup>[1]</sup>所以在Shafer理论中，不但有两个框架相容的概念，而且还有一个框架族相容的概念。

两个框架当其中的一个是另一个的精细时，定义这两个框架相容。一般情况下，两个框架当不满足其中的一个是另一个的精细时也可以定义相容，只要要求基于这两个框架的特性和概念是相容的，也即这两个框架必须具有一个公共的精细化框架。据此又可定义一个相容的框架族为两两相容的框架的汇集。

相容框架族的数学结构如下。

设 $\mathcal{F}$ 是有限非空集合的非空汇集， $\mathcal{F}$ 中的每两个集合之间都没有任何公共元素。 $\mathcal{R}$ 是映射的非空汇集，其中每一映射 $\omega \in \mathcal{R}$ 具有形式 $\omega: 2^{\Theta_1} \rightarrow 2^{\Theta_2}$  ( $\Theta_1, \Theta_2 \in \mathcal{F}$ ) 且是一个细分。那么 $\mathcal{F}$ 是一个具有细分 $\mathcal{R}$ 的相容的识别框架族， $\mathcal{F}, \mathcal{R}$ 必须满足

(1) 细分的合成：如果 $\omega_1: 2^{\Theta_1} \rightarrow 2^{\Theta_2}$  和  $\omega_2: 2^{\Theta_2} \rightarrow 2^{\Theta_3}$  在 $\mathcal{R}$ 中，那么 $\omega_2 \circ \omega_1$ 在 $\mathcal{R}$ 中，

(2) 粗化的恒等性：如果 $\omega_1: 2^{\Theta_1} \rightarrow 2^{\Theta_2}$  和  $\omega_2: 2^{\Theta_2} \rightarrow 2^{\Theta_2}$  在 $\mathcal{R}$ 中，且满足： $\forall \theta_1 \in \Theta_1, \exists \theta_2 \in \Theta_2$  使得  $\omega_1(\{\theta_1\}) = \omega_2(\{\theta_2\})$ ，则 $\Theta_1 = \Theta_2$  且  $\omega_1 = \omega_2$ ；

(3) 细分的恒等性：如果 $\omega_1: 2^{\Theta} \rightarrow 2^{\Theta}$  和  $\omega_2: 2^{\Theta} \rightarrow 2^{\Theta}$  都在 $\mathcal{R}$ 中，则 $\omega_1 = \omega_2$ ；

(4) 粗化的存在性：设想 $\Omega \in \mathcal{F}$ ，且  $A_1, A_2, \dots, A_n$  是 $\Omega$ 的划分，那么在 $\mathcal{F}$ 中存在一个粗化框架对应于该划分。

即在 $\mathcal{A}$ 中存在一个细分 $\omega: 2^0 \rightarrow 2^0$ , 使得 $\forall i, i=1, \dots, n$ ,  
 $\exists \theta \in \Theta$  具有 $\omega(\{\theta\}) = A_i$ ;

(5) 细分的存在性: 如果 $\Theta \in \mathcal{F}$ ,  $\theta \in \Theta$ ,  $n$ 是正整数,  
 那么存在细分 $\omega: 2^0 \rightarrow 2^0 (\in \mathcal{A})$ , 使得 $\omega(\{\theta\})$ 具有 $n$ 个元素;

(6) 公共精细的存在性: 如果 $\Theta_1, \Theta_2 \in \mathcal{F}$ , 那么存在  
 $\mathcal{A}$ 中的细分 $\omega_1: 2^{0_1} \rightarrow 2^{0_1}$  和 $\omega_2: 2^{0_2} \rightarrow 2^{0_2}$  使得 $\Omega_1 = \Omega_2$ 。

总的来讲, 当我们讨论一个框架时, 该框架可以被放在一个满足相容性定义的框架族中; 一般情况下, 当我们讨论细分或者粗化或者相容框架时, 问题中的所有框架都可以一块被放在这样一个框架族中。

不过, 为了无限细分时它的每一个框架的可允许性而设计的相容框架数学结构的条件(5)说明这种无限细分已经在族 $\mathcal{F}$ 中得到, 所以整个相容框架族的概念只是作为一种潜在的可能性而引出的, 在实际应用中, 从不使用整个相容框架族。相容框架族只是在理论分析, 比如在表示一些条件时, 才起作用。

用相容框架族定义中的条件(6)我们可以推出, 对于相容框架的任意有限汇集 $\{\Theta_1, \dots, \Theta_n\}$ 来讲都存在一个公共的精细化框架, 将其中那个最粗的框架(即它为其他公共精细化框架的粗框架)称为 $\Theta_1, \dots, \Theta_n$ 的最小精细化框架, 记为 $\Theta_1 \otimes \dots \otimes \Theta_n$ 。

最小精细化框架具有许多好的性质, 限于篇幅, 在此不再讨论, 请参看[1]。



### §3 相容的信度函数

设 $\Theta_1$ 和 $\Theta_2$ 是两个相容的识别框架, 信度函数 $\text{Bel}_1$ 和 $\text{Bel}_2$ 分别定义在 $\Theta_1$ 和 $\Theta_2$ 上。显然, 只要在由 $\Theta_1$ 和 $\Theta_2$ 共同识别的命题上 $\text{Bel}_1$ 与 $\text{Bel}_2$ 相等, 那么 $\text{Bel}_1$ 与 $\text{Bel}_2$ 就可以表示相同的判决。在这种情况下, 称 $\text{Bel}_1$ 与 $\text{Bel}_2$ 相容。

判断一个命题是否由 $\Theta_1$ 和 $\Theta_2$ 共同识别, 可求助于 $\Theta_1, \Theta_2$ 的最小精细化框架 $\Theta_1 \otimes \Theta_2$ 和细分

$$\omega_1: 2^{\Theta_1} \rightarrow 2^{\Theta_1 \otimes \Theta_2}; \quad \omega_2: 2^{\Theta_2} \rightarrow 2^{\Theta_1 \otimes \Theta_2}$$

对于 $A_1 \subset \Theta_1$ 和 $A_2 \subset \Theta_2$ , 当 $\omega_1(A_1) = \omega_2(A_2)$ 时,  $A_1$ 和 $A_2$ 代表同样的命题, 而该命题可以被 $\Theta_1$ 和 $\Theta_2$ 共同识别。所以 $\text{Bel}_1$ 与 $\text{Bel}_2$ 相容意味着当 $A_1 \subset \Theta_1, A_2 \subset \Theta_2$ 且 $\omega_1(A_1) = \omega_2(A_2)$ 时,  $\text{Bel}_1(A_1) = \text{Bel}_2(A_2)$ 。

设框架 $\mathcal{Q}$ 是框架 $\Theta$ 的一个精细框架,  $\omega: 2^{\Theta} \rightarrow 2^{\mathcal{Q}}$ 是一个细分,  $\text{Bel}: 2^{\Theta} \rightarrow [0, 1]$ 是 $\Theta$ 上的一个信度函数,  $\text{Bel}_0: 2^{\Theta} \rightarrow [0, 1]$ 是 $\Theta$ 上的一个集函数。可以证明: 只有当 $\text{Bel}_0$ 被定义成

$$\forall A \in 2^{\Theta} \quad \text{Bel}_0(A) = \text{Bel}(\omega(A))$$

时,  $\text{Bel}$ 与 $\text{Bel}_0$ 才是相容的 (事实上, 这样定义的 $\text{Bel}_0$ 也为一个信度函数)。此时称 $\text{Bel}_0$ 是 $\text{Bel}$ 在 $\Theta$ 上的限制, 用 $\text{Bel}|_{2^{\Theta}}$ 来表示。

从定义中可以看出, 若将 $2^{\Theta}$ 看成 $2^{\mathcal{Q}}$ 的子集, 那么 $\text{Bel}_0$ 的确是 $\text{Bel}$ 在 $2^{\Theta}$ 上的限制。

**定理4.3.1** 设 $\omega: 2^{\Theta} \rightarrow 2^{\mathcal{Q}}$ 是一个细分,  $\text{Bel}: 2^{\Theta} \rightarrow [0, 1]$ 是一个信度函数。令 $\mathcal{C}$ 和 $\mathcal{C}_0$ 分别表示 $\text{Bel}$ 和 $\text{Bel}|_{2^{\Theta}}$ 的核心,

$m$ 和 $Q$ 分别表示 $\text{Bel}$ 的基本可信度分配和众信度函数,  $m_0$ 和 $Q_0$ 分别表示 $\text{Bel}|2^\Theta$ 的对应函数,  $\bar{\theta}$ 表示 $\omega$ 的外收缩, 则

$$\mathcal{C}_0 = \bar{\theta}(\mathcal{C})$$

$$\forall A \subset \Theta$$

$$m_0(A) = \sum_{\substack{B \subset \Theta \\ A = \bar{\theta}(B)}} m(B)$$

$$Q_0(A) = \sum_{\substack{B \subset \Theta \\ A \subset \bar{\theta}(B)}} q(B)$$

$$\text{特别是 } \forall A \subset \Theta \quad m_0(A) \geq m(\omega(A)) \\ Q_0(A) \geq Q(\omega(A))$$

该定理告诉我们如何去计算一个信度函数的限制的基本可信度分配和众信度函数。

该定理的最后两式反映了对于一个命题, 它的基本可信数和众信数在框架转化过程中是递减的而不是递增的。

## §4 相容框架的独立性

两个相容的框架可以分为独立的与不独立的。如果对两个相容的框架而言, 由一个框架识别的命题不能非平凡地(non-trivially)蕴含由另一框架所识别的命题, 则称这两个框架是独立的。

要想知道由框架 $\Theta_1$ 识别的命题是否蕴含由另一相容框架 $\Theta_2$ 识别的命题, 与上相同我们必须求助于一个能够识别两个命题的框架, 而且验证在那个框架里第一个命题对应的

子集是否包含于第二个命题对应的子集。为简单起见，仍然考虑 $\Theta_1 \otimes \Theta_2$ 。因此当 $\Theta_1$ 和 $\Theta_2$ 相容时说对应于 $A_1 \subset \Theta_1$ 的命题蕴含对应于 $A_2 \subset \Theta_2$ 的命题就是说 $\omega_1(A_1) \subset \omega_2(A_2)$ 。其中： $\omega_1: 2^{\Theta_1} \rightarrow 2^{\Theta_1 \otimes \Theta_2}$ 和 $\omega_2: 2^{\Theta_2} \rightarrow 2^{\Theta_1 \otimes \Theta_2}$ 是 $\Theta_1$ 、 $\Theta_2$ 到最小精细化框架 $\Theta_1 \otimes \Theta_2$ 的细分。如果 $A_1 = \emptyset$ 或 $A_2 = \Theta_2$ ，则称该蕴含是平凡的。由此不难写出独立框架的数学定义。

**定义4.4.1** 设 $\Theta_1$ 和 $\Theta_2$ 是相容框架， $\omega_1: 2^{\Theta_1} \rightarrow 2^{\Theta_1 \otimes \Theta_2}$ ， $\omega_2: 2^{\Theta_2} \rightarrow 2^{\Theta_1 \otimes \Theta_2}$ 分别是 $\Theta_1$ 、 $\Theta_2$ 到它们的最小精细化框架的细分，如果当 $A_1 \subset \Theta_1$ ， $A_2 \subset \Theta_2$ ，而且 $\omega_1(A_1) \subset \omega_2(A_2)$ 时有 $A_1 = \emptyset$ 或 $A_2 = \Theta_2$ ，则称 $\Theta_1$ 和 $\Theta_2$ 独立。

类似地还可以定义多个相容框架的独立性。限于篇幅，不再讨论。

## §5 框架的收缩与扩张

收缩与扩张的概念和精细与粗化的概念是完全不同的。

如果在一个框架 $\Theta_2$ 上加上一个新假设，则 $\Theta_2$ 的元素必定要减少，设 $\Theta_1$ 是由 $\Theta_2$ 加上一个新假设而得到的识别框架，则 $\Theta_1$ 作为可能事件或结果的一个集合将是可能结果 $\Theta_2$ 的一个子集。此时，我们称 $\Theta_1$ 是 $\Theta_2$ 的一个收缩， $\Theta_2$ 是 $\Theta_1$ 的一个扩张。

值得注意的是，框架的收缩与扩张之间的关系与粗化和精细之间的关系是非常不同的。如果 $\Theta$ 是 $\Omega$ 的粗化，那么 $\Theta$ 和 $\Omega$ 是相容的，这两个框架是基于相同的概念和假设的，所以当 $\Theta$ 的一个子集对应于 $\Omega$ 的一个子集时，这两个子集所对应的命题是完全相同的。但是，如果 $\Theta_1$ 是 $\Theta_2$ 的一个收

缩，那么 $\Theta_1$ 和 $\Theta_2$ 是不相容的。 $\Theta_1$ 与 $\Theta_2$ 基于不同的假设，因此 $\Theta_1$ 的一个给定的子集所对应的命题与大框架 $\Theta_2$ 中对应子集多少具有一些不同的意义。

那么实践中究竟采取什么样的假设，用什么作为我们的识别框架？这主要取决于问题的不同以及我们对讨论的问题所具有的知识的多少。在这次研究中取定的框架，在下次研究中由于我们的知识的增加以及研究问题的角度的改变可以收缩或扩张该框架。在本次研究中，一旦取定了框架，那么我们的所有理论就都是可使用的。Shafer信度理论适合于取定了框架的任一种似真推理。

## 第五章 信度函数的分类 及证据的权重

前几章讨论了有关信度函数的问题，信度函数是我们根据证据对各种可能性进行判决的结果，是证据在一个识别框架上的体现。本章将从另一角度针对各种特殊类型的信度函数讨论证据，引出描述证据强弱程度的量——证据的权重。

### §1 简单支持函数

一批证据在一个框架 $\Theta$ 上的作用通常是支持许多被 $\Theta$ 所识别的命题，但是程度不同。最简单的情况就是该批证据精确地，毫不含混地支持一个 $\Theta$ 的单个非空子集 $A$ 。此时我们说证据的作用局限于为 $A$ 提供一个确定的支持度。证据以一定的程度支持 $A$ ，当然它也以同样的程度支持被 $A$ 所蕴含的命题。但是对其他 $\Theta$ 可识别的命题却不提供任何支持。

设 $s$ 是对 $A$ 的支持度 ( $0 \leq s \leq 1$ )，则 $\forall B \subset \Theta$ ，证据所提供的支持度为

$$S(B) = \begin{cases} 0 & A \not\subset B \\ s & A \subset B, B \neq \Theta \\ 1 & B = \Theta \end{cases} \quad (5.1.1)$$

由此所定义的函数  $S: 2^{\Theta} \rightarrow [0, 1]$  称为简单支持函数。

显然, 一个简单支持函数是一个信度函数。事实上, 针对  $A$  的简单支持函数  $S: 2^{\Theta} \rightarrow [0, 1]$  所对应的基本可信度分配为

$$\forall B \subset \Theta \quad m(B) = \begin{cases} S(A) & B = A \\ 1 - S(A) & B = \Theta \\ 0 & \text{其他} \end{cases} \quad (5.1.2)$$

对于两个针对同一集合  $A$  的简单支持函数  $S_1$  和  $S_2$ , 其合成可以采用贝努力合成法则。该法则可以陈述为: 假设证据  $E_1$  对  $A$  的支持度为  $s_1$ , 证据  $E_2$  对  $A$  的支持度为  $s_2$ , 则  $E_1$  与  $E_2$  的合成对  $A$  的支持度为  $s_1 + s_2 - s_1 \cdot s_2$ 。这个法则是 Dempster 合成法则的特例。在这种情况下, 证据  $E_1$  和  $E_2$  的合成称为同类证据 (Homogeneous Evidence)。此时,  $E_2$  的加入必定增强  $E_1$ 。

设证据  $E_1$  精确地、毫不含混地支持集合  $A$ , 证据  $E_2$  精确地、毫不含混地支持集合  $B$ ,  $A \neq B$ 。则证据  $E_1$  和  $E_2$  的合成仅为不同类证据 (Heterogeneous Evidence)。此时,  $E_2$  的对  $E_1$  来讲可以加强, 也可以减弱。

对于不同类证据  $E_1$  和  $E_2$  产生的简单支持函数  $S_1$  和  $S_2$ , 该如何合成呢?

设  $A \cap B \neq \emptyset$  且  $A$  与  $B$  互不包含,  $S_1(A) = s_1$ ,  $S_2(B) = s_2$ , 则由 Dempster 合成法则可知,  $S_1$  与  $S_2$  的合成是如下的函数

$$VC \subseteq \Theta \quad S(C) = \begin{cases} 0 & C \not\supset A \cap B \\ s_1 \cdot s_2 & C \supset A \cap B, C \not\supset A, C \not\supset B \\ s_1 & C \supset A, C \not\supset B \\ s_2 & C \supset B, C \not\supset A \\ 1 - (1 - s_1)(1 - s_2) & C \supset A, C \supset B, C \neq \Theta \\ 1 & C = \Theta \end{cases}$$

我们看到，对被  $A$  所蕴含而不能被  $B$  蕴含的命题提供  $s_1$  的支持；对被  $B$  所蕴含而不能被  $A$  蕴含的命题提供  $s_2$  的支持。这说明，对于  $E_1$  来讲，证据  $E_2$  的加入对被  $A$  所蕴含而不能被  $B$  所蕴含的命题特别是对  $A$  本身无任何影响；对  $E_2$  来讲也有同样的情况。而对可被  $A, B$  的共同部分所蕴含的命题特别是  $A \cap B$  提供  $s_1, s_2$  的支持，对不能被  $A \cap B$  蕴含的命题不提供任何支持。这些都符合我们的直观理解。

当  $A \subset B$  时， $E_1$  与  $E_2$  不是真正的不同类， $E_1$  与  $E_2$  在一定程度上具有一些共同的特性， $E_1$  与  $E_2$  的合成可用下列函数来表示

$$VC \subseteq \Theta \quad S(C) = \begin{cases} 0 & C \not\supset A \\ s_1 & C \supset A, C \not\supset B \\ 1 - (1 - s_1)(1 - s_2) & C \supset B, C \neq \Theta \\ 1 & C = \Theta \end{cases}$$

可见，对仅能由  $A$  蕴含不能由  $B$  蕴含的命题， $E_2$  的加入对其信度不产生任何影响；另外，对不能由  $A$  蕴含的命题， $E_2$  的加入对其信度也不产生任何影响。

当  $A \cap B = \phi$  时，情况变得复杂起来，此时  $E_1$  与  $E_2$  说的完全是两件不同的事。 $E_1$  与  $E_2$  不仅仅是不同类的，而且还是矛

盾的。其中一方的作用都要被另一方所消弱。 $E_1$ 与 $E_2$ 产生的简单支持函数 $S_1$ 和 $S_2$ 可合成为：

$$\forall C \subset \Theta$$

$$S(C) = \begin{cases} 0 & C \supset A, C \supset B \\ \frac{s_1(1-s_2)}{1-s_1s_2} & C \supset A, C \not\supset B \\ \frac{s_2(1-s_1)}{1-s_1s_2} & C \supset B, C \not\supset A \\ \frac{s_1(1-s_2) + s_2(1-s_1)}{1-s_1s_2} & C \supset A, C \supset B, C \neq \Theta \\ 1 & C = \Theta \end{cases}$$

可见，由于 $E_2$ 的加入，使我们对 $A$ 的信度缩小了 $(1-s_2)/(1-s_1s_2)$ 倍，同样由于 $E_1$ 的作用，使得我们对 $B$ 的信度缩小了 $(1-s_1)/(1-s_1s_2)$ 倍。

以上讨论了两个简单支持函数的合成，多个简单支持函数的合成情况怎样呢？

## §2 可分离支持函数

可分离支持函数是简单支持函数或者两个以及两个以上简单支持函数的直和。尽管可分离支持函数不能包括所有适合于证据表示的信度函数，但对于能分解成同类证据的那些证据来讲，用可分离支持函数来表示是合适的。

设 $S$ 是一个可分离支持函数，则一定存在简单支持函数 $S_1, S_2, \dots, S_n$ 使 $S = S_1 \oplus S_2 \oplus \dots \oplus S_n$ 。但是，这种分解却不是唯一的。

**定理5.2.1** 设 $S$ 是一个 $\Theta$ 上的可分离支持函数， $S = S_1 \oplus$



$\cdots \oplus S_n$ , 此处  $S_i (i=1, \cdots, n)$  是简单支持函数。令  $\mathcal{C}$  表示  $S$  的核心,  $A_1, \cdots, A_n$  表示  $S_1, \cdots, S_n$  的除  $\Theta$  以外的焦元, 则

(1) 设  $S_{n+1}$  是空信度函数, 那么  $S_{n+1}$  是简单支持函数, 并且  $S = S_1 \oplus \cdots \oplus S_{n+1}$ ;

(2) 如果  $A_1 = A_2$ , 那么  $S_1 \oplus S_2$  是一个简单支持函数, 而  $S = (S_1 \oplus S_2) \oplus \cdots \oplus S_n$ ;

(3) 设  $S_{n+1}$  是一个简单支持函数, 除  $\Theta$  以外的焦元  $A_{n+1}$  满足  $S_{n+1}(A_{n+1}) < 1$  并且  $A_{n+1} \subset \mathcal{C}$ , 则  $S = S_1 \oplus \cdots \oplus S_{n+1}$ ;

(4) 设所有  $A_i$  都与  $\mathcal{C}$  具有非空交, 对每一  $i$ , 令  $S'_i$  表示针对  $A_i \cap \mathcal{C}$  且  $S'_i(A_i \cap \mathcal{C}) = S_i(A_i)$  的简单支持函数, 则  $S = S'_1 \oplus \cdots \oplus S'_n$ 。

根据定理 5.2.1, 对一个可分离支持函数, 如果我们要求所分解成的简单支持函数是非空的, 所有  $\Theta$  以外的焦元不等, 并且都包含在该可分离支持函数的核心中, 则这种分解就是唯一的。

**定理 5.2.2** 设  $S$  是一个非空可分离支持函数,  $\mathcal{C}$  为其核心, 则存在一个唯一的非空简单支持函数序列  $S_1, \cdots, S_n$  满足

(1)  $n \geq 1$ ;

(2)  $S = S_1 \oplus \cdots \oplus S_n$ ;

(3) 每个  $S_i$  的  $\Theta$  以外的焦元包含在  $\mathcal{C}$  中;

(4) 如果  $i \neq j$ , 那么  $S_i$  的焦元不等于  $S_j$  的焦元。

$S$  的这种唯一的分解  $S_1, \cdots, S_n$  称为正规分解 (Canonical decomposition)。  $S$  的正规分解可以通过  $S$  的任意一种分解得到。由定理 5.2.1, 得到一个任意分解以后, 首先去掉空函数, 再去掉焦元与核心不交的函数, 缩小与核心相交的焦元为该焦元与核心的交, 再合成具有公共焦元的函数, 最后

得到的这一系列简单支持函数就是 $S$ 的正规分解。

### §3 支持函数

支持函数(support function)是一类由可分离支持函数通过框架粗化得到的信度函数。这一类函数构成了适合于证据表示的信度函数的子类。

**定义5.3.1** 设 $Bel: 2^\Theta \rightarrow [0, 1]$ 是 $\Theta$ 上的信度函数, 如果存在 $\Theta$ 的精细框架 $\Omega$ 和 $\Omega$ 上的一个可分离支持函数 $S: 2^\Omega \rightarrow [0, 1]$ 使得 $Bel = S \upharpoonright 2^\Theta$ , 则称 $Bel$ 为支持函数。

由此定义可以看出, 可分离支持函数是支持函数。由此我们得到下面的一系列包含关系

$$\left\{ \begin{array}{c} \text{简单支持} \\ \text{函数} \end{array} \right\} \subset \left\{ \begin{array}{c} \text{可分离} \\ \text{支持函数} \end{array} \right\} \subset \left\{ \begin{array}{c} \text{支持} \\ \text{函数} \end{array} \right\} \subset \left\{ \begin{array}{c} \text{信度} \\ \text{函数} \end{array} \right\}$$

由前两节可知, 并非所有可分离支持函数都是简单的, 即存在不是简单支持函数的可分离支持函数。下面两定理说明后面的两个包含关系也是真包含关系。

**定理5.3.1** 设 $Bel$ 是信度函数,  $\mathcal{C}$ 是它的核心, 那么下列论断都是等价的

- (1)  $Bel$ 是支持函数;
- (2)  $\mathcal{C}$ 具有正的众信数;
- (3)  $\mathcal{C}$ 具有正的基本可信数。

**定理5.3.2** 设 $S$ 是一个可分离支持函数,  $A$ 和 $B$ 是它的两个焦元, 若 $A \cap B \neq \emptyset$ , 那么 $A \cap B$ 也是 $S$ 的焦元。

定理5.3.1说明, 例2.3.1中的信度函数不是一个支持函数, 这就是说最后的包含关系是真包含。下面根据定理

5.3.1和定理 5.3.2 构造一个不是可分离支持函数的支持函数，从而说明第二个包含关系也是真包含。

例5.3.1 令 $\Theta = \{a, b, c\}$ ，定义在 $\Theta$ 上的一个信度函数 Bel具有焦元 $\{a, b\}, \{a, c\}, \Theta$ ，且基本可信数全部为 $1/3$ 。由于 $\Theta$ 是焦元，所以由定理5.3.1知Bel是支持函数。但是，由于 $\{a, b\} \cap \{a, c\} = \{a\}$ 不是焦元，因而由定理5.3.2知，Bel不是可分离支持函数。

定理5.3.1列出了支持函数的几个等价条件，这几个等价条件当然也可以作为支持函数的定义。

不是支持函数的信度函数在下几节还要说明。在那里可以看到，这种类型的信度函数不能作为实际证据的表示。所以说，支持函数是适合于表示有限框架上的证据的信度函数的最大子类。

对于支持函数还有所谓的空扩张的概念。

定义5.3.2 设 $S_0: 2^\Theta \rightarrow [0, 1]$ 是一个支持函数， $\omega: 2^\Theta \rightarrow 2^\Theta$ 是一个细分，则由 $S(A) = \max_{\substack{B \subseteq \Theta \\ \omega(B) = A}} S_0(B)$ 定义的  $S: 2^\Theta \rightarrow$

$[0, 1]$ 是一个支持函数且满足 $S_0 = S|_{2^\Theta}$ ， $S$ 称为 $S_0$ 在 $\mathcal{Q}$ 上的空扩张(Vacuous Extension)，也称为 $S$ 被粗框架 $\Theta$ 所承载。

空扩张在估价(evaluation)证据时经常被提到。给定一批固定证据 $\mathcal{E}$ 和一个识别框架 $\Theta$ ，我们通常可以找到 $\Theta$ 的一个精细框架 $\mathcal{Q}$ 使 $\mathcal{E}$ 对 $\mathcal{Q}$ 无任何新的作用。但是，我们并不能保证对 $\Theta$ 的所有精细框架 $\mathcal{Q}$ 证据 $\mathcal{E}$ 都没有新的作用。事实上，对于一个单个的固定的识别框架来说，有时是根本不可能用尽一批给定证据的作用的。这个问题下面还要讨论。

对支持函数，还有许多理论问题可加以讨论，但限于篇

幅,在此不再进行,读者可参看[1]。

## §4 证据的权重

证据权重概念的引入也是Shafer对于证据理论的主要贡献之一。

前面说过,一批证据在一个框架 $\Theta$ 上的作用,最简单的情况就是以一定的程度精确地、毫不含混地支持 $\Theta$ 的一个给定的子集 $A$ 。很显然对 $A$ 的支持度 $s$ 的大小反映了该证据的有力程度。 $s$ 越大,证据越强; $s$ 越小,证据越弱。反之,证据的强弱程度也应该完全决定那批证据对 $A$ 的支持度 $s$ 。证据越强, $s$ 越大;证据越弱, $s$ 越小。反映证据强弱程度的一个量我们称为证据的权重(Weight of evidence)。由上分析可知,在支持度 $s$ 与证据的权重之间似乎存在着某种函数关系。事实上,这种函数在作了一些假定以后,可以明确地表示出来。

设指向 $A$ 的证据的权重为 $w$ ,则 $w$ 可取任何非负值,但是对 $A$ 的支持度 $s$ 必须在 $[0, 1]$ 之内,所以必须要求这样一个函数 $g: [0, \infty] \rightarrow [0, 1]$ 满足 $s = g(w)$ 且是连续的(这样的要求从直观上讲是合理的)。

设 $S_1$ 和 $S_2$ 是同一识别框架 $\Theta$ 上的两个支持函数,焦点均为 $A \subset \Theta$ ,  $S_1(A) = s_1$ ,  $S_2(A) = s_2$ 。又设 $w_1$ 和 $w_2$ 分别为 $S_1$ 和 $S_2$ 对应证据的权重,  $g(w_1) = s_1$ ,  $g(w_2) = s_2$ 。从直观上看权重的合成应该具有可加性,所以 $S_1$ 和 $S_2$ 的合成应该具有权重 $w_1 + w_2$ 。

由贝努力合成法则知: $S_1$ 与 $S_2$ 的合成对 $A$ 的支持度为 $s_1 +$

$s_2 - s_1 s_2 = 1 - (1 - s_1)(1 - s_2)$ 。所以, 得到

$$g(w_1 + w_2) = 1 - (1 - s_1)(1 - s_2)$$

令  $f(w) = 1 - g(w)$  则上关系变为

$$f(w_1 + w_2) = f(w_1) \cdot f(w_2)$$

因此,  $f$  必须为指数函数设为  $f = e^{cw}$ , 其中  $c$  为常数。为使  $f$  将  $[0, \infty]$  映到  $[0, 1]$ , 必须要求  $c$  为负值, 再考虑到权重的度量是相对的, 所以, 为方便起见, 令  $c = -1$ 。这样就得到

$$f(w) = e^{-w}$$

而  $g(w) = 1 - e^{-w}$

即  $s = 1 - e^{-w}$  (5.4.1)

解出  $w$  为  $w = -\ln(1 - s)$  (5.4.2)

从而可计算要产生这么大的支持度需要多么大的权重的证据。同样可知有这么大权重的证据可产生多么大的支持度。

权重可取从0到无限大之间的所有值。当权重为0时, 产生的支持度也为0, 即对给定的命题不提供任何支持; 当权重为无限大时, 产生的支持度为1, 即对给定命题提供百分之百的支持, 在给定证据下所给命题成为确定性命题。

以上对简单支持函数讨论了证据的权重的概念。事实上, 不管证据在识别框架上产生什么类型的函数, 它都有一个权重的概念。先看可分离支持函数的情况。

一个可以写为一些简单支持函数直和的可分离支持函数可以直接用基于那些简单支持函数的证据的权重来描述。

设  $S = S_1 \oplus \dots \oplus S_n$ ,  $S$  所对应的众信函数为  $Q$ 。  $S_i (i = 1, \dots, n)$  是针对  $\Theta$  的非空真子集  $A_i$  的简单支持函数,  $Q_i, w_i$  分别为  $S_i$  所对应的众信函数和权重。

$\therefore S_i$  的基本可信度分配为

$$\begin{cases} m_i(A_i) = s_i = 1 - e^{-w_i} \\ m_i(\Theta) = 1 - s_i = e^{-w_i} \end{cases}$$

$$\therefore Q_i(A) = \begin{cases} 1 & \text{如果 } A \subset A_i \\ e^{-w_i} & \text{如果 } A \not\subset A_i \end{cases}$$

由 Dempster 合成法则的众信函数形式知

$$\begin{aligned} Q(A) &= K \prod_{i=1}^n Q_i(A) = K \prod_{A \not\subset A_i} e^{-w_i} \\ &= K \exp\left(-\sum_{A \not\subset A_i} w_i\right) \end{aligned} \quad (5.4.3)$$

$$\text{其中 } K = \left( \sum_{\substack{B \subset \Theta \\ B \neq \emptyset}} (-1)^{|B|-1} \exp\left(-\sum_{A_i \in B} w_i\right) \right)^{-1}$$

下面的定理对于进一步讨论问题是非常有用的。

**定理 5.4.1** 设  $S_1, \dots, S_n$  为简单支持函数，这些简单支持函数分别针对着  $\Theta$  的非空真子集  $A_1, \dots, A_n$ ，对应的证据的权重分别为  $w_1, \dots, w_n$ ，则  $S_1 \oplus \dots \oplus S_n$  存在的充分必要条件为

$$\mathcal{C} = \bigcap_{i=1}^n A_i \neq \emptyset \quad (5.4.4)$$

(当所有  $w_i$  有限时  $\mathcal{C} \triangleq \Theta$ )

如果  $S_1 \oplus \dots \oplus S_n$  存在，那么核心为  $\mathcal{C}$ ，众信度  $Q$  由 (5.4.3) 式定义，而且  $Q(A) > 0$  的充分必要条件为  $A \subset \mathcal{C}$  (对任意一个信度函数在它的核心上众信数不一定大于 0)。

根据该定理即可定义证据的估价 (Assessment of

Evidence)。

**定义5.4.1** 设权重为 $w_1, \dots, w_n$ 的证据所对应的集合为 $A_1, \dots, A_n$ , 且满足 $\mathcal{E} = \bigcap_{w_i > 0} A_i \neq \emptyset$ , 那么由下式定义的函数 $W: 2^\Theta \rightarrow [0, \infty]$ 称为和各 $w_i$ 相联系的证据的估价

$$W(A) = \begin{cases} \sum_{A_i \cap \mathcal{E} = A} w_i & \text{如果 } A \subset \mathcal{E}, A \neq \emptyset \text{ 且 } A \neq \mathcal{E} \\ \infty & \text{如果 } A = \mathcal{E} \\ 0 & \text{如果 } A = \emptyset \text{ 或 } A \not\subset \mathcal{E} \end{cases} \quad (5.4.5)$$

这个定义说明对于可分离支持函数的一种分解总可以定义一个证据的估价。

另外, 由该定义也可以看出, 当 $A$ 不是相应的可分离支持函数之焦元时 $W(A) = 0$ , 且若 $W(B) > 0$ , 估价 $W(B)$ 一方面说明证据对焦元 $B$ 的某种支持程度, 另一方面也刻划了对不在 $B$ 中的 $\Theta$ 的其他点的怀疑程度。

**定理5.4.2** 设权重为 $w_1, \dots, w_n$ 的证据所针对的集合为 $A_1, \dots, A_n$ ,  $\mathcal{E} = \bigcap_{w_i > 0} A_i \neq \emptyset$ ,  $Q$ 是由这些权重所决定的众信函数,  $W$ 是和这些 $w_i$ 相联系的证据的估价。则

(1)  $Q$ 可被 $W$ 决定

$$\forall A \subset \Theta \quad A \neq \emptyset$$

$$Q(A) = K \exp\left(-\sum_{\substack{B \subset \Theta \\ A \cap B}} W(B)\right) \quad (5.4.6)$$

$$\text{其中} \quad K = \left( \sum_{\substack{B \subset \Theta \\ B \neq \emptyset}} (-1)^{|B|+1} \exp\left(-\sum_{\substack{C \subset \Theta \\ B \subset C}} W(C)\right) \right)^{-1}$$

(2)  $W$  可从  $Q$  中得到

$$W(A) = \begin{cases} \sum_{\substack{B \subseteq \mathcal{E} \\ A=B}} (-1)^{|B-A|} I_{\pi} Q(B) & \text{如果 } A \subseteq \mathcal{E}, A \neq \phi, A \neq \mathcal{E} \\ \infty & \text{如果 } A = \mathcal{E} \\ 0 & \text{如果 } A = \phi \text{ 或 } A \not\subseteq \mathcal{E} \end{cases} \quad (5.4.7)$$

该定理中的(2)说明对于一个可分离支持函数  $S$ ，不管其分解如何，所定义的证据的估价是唯一的；定理中的(1)说明一个证据的估价可以唯一地决定一个可分离支持函数，这正符合证据与由它所产生的信度之间的直观的对应关系，这说明用可分离支持函数  $S$  的证据的估价去估计产生  $S$  的证据是合适的；证据的估价可以作为证据的权重的推广。

上面说过，一个简单支持函数是基于同类证据的。对于不是同类证据的证据(不同类证据 heterogeneous evidence)内部多少要表现出一些冲突，如何反映证据内部的冲突呢？对于产生可分离支持函数的证据，可定义内部冲突的权重 (Weight of internal Conflict)。

**定义 5.4.2** 一个可分离支持函数  $S$  内部冲突的权重定义为

当  $S$  是简单支持函数时是 0；

当  $S$  不是简单支持函数时是对于各种可能的分解  $S = S_1 \oplus \dots \oplus S_n$  ( $S_i$  为简单支持函数)，数量  $\text{Con}(S_1, \dots, S_n)$  的下确界。

内部冲突权重用符号  $W$  加一个合适的下标来表示，当注意到  $S$  本身时写成  $W_s$ ，当注意到  $Q$  时写成  $W_Q$ 。



$W$ 可以根据表示 $S$ 的证据的估价,  $S$ 的众信度函数以及 $S$ 的各种分解求出。计算公式可参见[1], 在此不再给出。

**定义5.4.3** 由

$$V(A) = \begin{cases} \sum_{\substack{B \in \Theta \\ A \cap B \neq \emptyset}} W(B) & A \neq \phi \\ 0 & A = \phi \end{cases} \quad (5.4.8)$$

定义的函数  $V: 2^\Theta \rightarrow [0, \infty)$  称为冲突函数 (impingement function)。

由于每个权重  $W(B)$  怀疑所有不在它的焦元  $B$  中的  $\Theta$  的点, 因而只要  $A \cap B \neq \emptyset$ ,  $W(B)$  就怀疑  $A$  的一个部分,  $V(A)$  是至少怀疑  $A$  的一个部分的总权重, 即  $V(A)$  是在  $A$  上发生冲突的证据的总权重。

有了冲突函数  $V$  以后, 根据 (5.4.6) 式可以得到

$$Q(A) = K e^{-V(A)} \quad (5.4.9)$$

其中 
$$K = \left( \sum_{\substack{B \in \Theta \\ B \neq \emptyset}} (-1)^{|B|+1} e^{-V(B)} \right)^{-1}$$

将 (5.4.7) 式与 (5.4.9) 式进行组合可得

$$W(A) = - \sum_{\substack{B \in \Theta \\ A \cap B \neq \emptyset}} (-1)^{|B|-|A|} V(B) \quad (5.4.10)$$

所以, 如果我们知道  $W$ , 我们可以通过 (5.4.8) 式找到  $V$ 。如果我们知道  $V$ , 那么可通过 (5.4.10) 式找到  $W$ 。因此冲突函数  $V$  与估价  $W$  所确切表示的信息相同。冲突函数  $V$  从另一侧面反映了产生可分离支持函数  $S$  的证据。冲突函数  $V$  也可以看成是证据权重的推广。

由以上分析可知, 对于产生可分离支持函数的证据来讲,

用可分离支持函数、证据的估价、冲突函数三者来描述都是可以的，如同一批同类证据可以由简单支持函数也可以由证据的权重来描述一样。

以上我们将证据权重的概念推广到了可分离支持函数，那么对于一般的支持函数可否定义与证据权重相类似的概念呢？

因为支持函数是通过可分离支持函数在一个粗框架上的限制来定义的，所以要研究支持函数就必须涉及一个相容框架族。又因为一批证据同时作用于相容框架的整个族，在该族的每一成员上决定一个支持函数，所以一批证据的作用不单单是产生一个支持函数，而是产生一个在第四章§3意义下相容的支持函数族。在这种情况下，要讨论产生一个支持函数的证据的权重或与此类似的概念似乎就是不可能的。但是，对一个框架 $\Theta$ 上的支持函数 $S$ 来讲，却可以讨论在 $\Theta$ 上产生支持函数 $S$ 的整个证据 $\mathcal{E}$ 被 $\Theta$ 所识别的那部分权重。

在[1]中Shafer指出了一种定义这种概念的可能方法。先看一个定义。

**定义5.4.4** 设 $\mathcal{F}$ 是一个相容识别框架族， $\mathcal{E}$ 是作用于 $\mathcal{F}$ 的一批证据，如果存在一个框架 $\Theta \in \mathcal{F}$ 承载了 $\Theta$ 的每一个精细框架 $\Omega \in \mathcal{F}$ 上的支持函数 $S_{\Omega}^{\mathcal{E}}$ ，则称为证据 $\mathcal{E}$ 准确地作用于 $\mathcal{F}$  ( $\mathcal{E}$  affects  $\mathcal{F}$  Sharply)；这样一个框架 $\Theta$ 称为用尽了 $\mathcal{E}$ 在 $\mathcal{F}$ 上的作用(exhaust the impact of  $\mathcal{E}$  on  $\mathcal{F}$ )。

当 $\Theta$ 用尽 $\mathcal{E}$ 在 $\mathcal{F}$ 上的作用时， $S_{\Omega}^{\mathcal{E}}$ 决定了整个族 $\{S_{\Omega}^{\mathcal{E}}\}_{\Omega \in \Theta}$ 。因为，如果 $\Theta$ 承载了 $\Theta$ 的每一个精细框架上的支持函数，那么在任何一个给定的框架 $\Omega \in \mathcal{F}$ 上的支持函数都是 $S_{\Omega}^{\mathcal{E}}$ 在 $\Theta \otimes \Omega$ 上的空扩张在 $\Omega$ 上的限制。但是如果 $\mathcal{E}$ 不是准确地作用于

$\mathcal{F}$ ，那么，在 $\mathcal{F}$ 中没有一个单个框架能够充分好的用尽 $\mathcal{E}$ 的所有作用，通过进一步细分，我们的框架总是可能进一步识别证据的作用。

很显然，处理准确地作用于 $\mathcal{F}$ 的证据是很容易的。事实上，在简单支持函数与可分离支持函数的讨论中，我们都默默地假定我们的框架可以用尽证据的作用，所以在那种情况下，讨论证据的权重是非常自然的。但是，当我们的框架不能用尽证据 $\mathcal{E}$ 的作用时，对于简单支持函数或可分离支持函数，上面所定义的证据的权重也只能看成是整个证据 $\mathcal{E}$ 被 $\Theta$ 所识别的那部分权重。

设想孤立地考虑一个支持函数 $S: 2^{\Theta} \rightarrow [0, 1]$ 。假定 $S$ 是作为一些 $\Theta$ 的精细框架 $\mathcal{Q}$ 上的一些可分离支持函数的限制而产生的，但是，我们不知道关于这个可分离支持函数或甚至关于 $\mathcal{Q}$ 的任何细节。在这种情况下，关于基于 $S$ 的证据的权重我们能说什么呢？

这个问题的一个自然反应就是去考虑集合

$$\mathcal{W}_S = \{W \mid W \text{ 是 } \Theta \text{ 的一些精细框架上的证据的估价, 而且由 } W \text{ 所决定的可分离支持函数 } S_W \text{ 满足 } S_W \mid 2^{\Theta} = S\}$$

去问在这个集合的所有元素上是否存在公共特性，即在所有可以产生 $S$ 的估价上是否存在公共特性。

Shafer没有解决这个问题，但是他给出了一些有用的事实

- (1)  $\mathcal{W}_S$ 总是包含一些权重是任意大的估价；
- (2)  $\mathcal{W}_S$ 包含在被 $\Theta$ 识别的命题上一点也不产生任何权重的估价。

(这两个事实有定理保证。)

事实(1)提醒我们应该寻找在一定意义下,为了决定 $S$ 所需的最小的证据的权重;但是事实(2)又表明不应该试图明确地针对对应于 $\Theta$ 的子集的权重的最小值。

Shafer指出:一个明显的方案就是去求冲突函数的最小值。<sup>[1]</sup>

设 $S: 2^\Theta \rightarrow [0, 1]$ 是一个支持函数,令

$$\mathcal{V}_S = \{V \mid V \text{ 是 } \Theta \text{ 的一些精细框架上的一个冲突函数} \\ \text{而且由 } V \text{ 决定的可分离支持函数满 } S_V \text{ 是 } S_V: 2^\Theta \\ = S\}$$

对每一 $V \in \mathcal{V}_S$ , 令 $\omega_V$  ( $V$ 定义在上面)表示 $\Theta$ 的一个精细框架,  $\omega_V: 2^\Theta \rightarrow 2^\Theta$ 是一个细分,  $W_V$ 是 $S_V$ 的内部冲突的权重。又令 $W_S = \inf_{V \in \mathcal{V}_S} W_V$

$$\forall A \subset \Theta \quad \underline{V}_S(A) = \inf_{V \in \mathcal{V}_S} V(\omega_V(A))$$

则数量 $W_S$ 可以称为被 $\Theta$ 识别的冲突的权重;函数 $\underline{V}_S: 2^\Theta \rightarrow [0, \infty)$ 可以称为由 $\Theta$ 识别的冲突函数。

对于可分离支持函数,在一定条件下

$$W_S = \underline{W}_S \quad \underline{V}_S = \underline{V}_S$$

这说明上两式对于在一般情况下证据权重的概念从可分离支持函数推广到支持函数来说,在一定程度上是成功的。

## §5 准支持函数

所谓准支持函数就是不是支持函数的信度函数,这类信度函数可以表示为一个可分离支持函数序列的极限或者该极

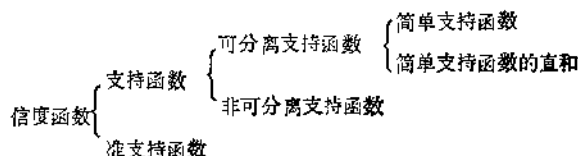
限的一个限制。准支持函数的产生需要一种无限矛盾权重的证据。所以，对于实际证据的表示，它是不合适的。

值得注意的是大部分贝叶斯信度函数（非平凡的贝叶斯信度函数，即 $\exists A \subset \Theta$ , 使 $\text{Bel}(A) > 0$ ,  $\text{Bel}(\bar{A}) > 0$ 而且 $\text{Bel}(A) + \text{Bel}(\bar{A}) = 1$ ）是准支持函数，即大多数贝叶斯信度函数不是支持函数。所以贝叶斯信度函数一般不适合于证据的表示。特别是把几率作为我们的信度而得到的贝叶斯信度函数不是支持函数，因而不适合于证据的表示。

相对于准支持函数还有一类对于证据表示非常方便的信度函数——一致信度函数。这类信度函数是可分离支持函数的一个子类，而且在信度理论与模糊数学和可能性理论的联系中起非常重要的作用。参见下章 § 5。

## §6 信度函数的分类

综上所述，信度函数的分类可表示为



尽管Dempster合成法则可用于所有满足信度函数的三个公理的集函数，但是不能认为所有信度函数在表示证据时都是有用的。可以由实际证据产生的信度函数仅仅是整个信度函数类的一个子集。简单支持函数与可分离支持函数是适合

于证据表示的信度函数类，一般的支持函数也可以用于表示证据，但是准支持函数却不能用于实际证据的表示。

## 第六章 一致支持函数

上章讲到，准支持函数不适合于证据的表示。相对于准支持函数，一致支持函数(Consonant support function)是一类特别适合于证据表示的函数。

一致支持函数是可分离支持函数的特例。在产生一致支持函数的证据中没有明显的冲突。虽然产生一致支持函数的证据不是同类证据，但是这类证据“指向一个单个方向”(pointing in a single direction)，只是在改变焦元精确性的意义下表现出一些不同类特性。这类证据并不是针对一个单个命题，而是针对一系列命题，这一系列命题满足后面的一个是前面的一个的结论。一致支持函数特别适合于这样的无明显冲突的证据的表示。

### §1 一致支持函数

**定义6.1.1** 如果一个信度函数的焦元是套形的(Nested)，即它的焦元可以顺序排列，以至于每一个可以包含在下一个中，则该信度函数称为一致支持函数。

一致支持函数具有许多等价定义。

**定理6.1.1** 设  $\text{Bel}: 2^\Theta \rightarrow [0,1]$  是信度函数,  $\text{pl}, Q$  是对应的似真度函数和众信度函数, 则下列论断等价

(1)  $\text{Bel}$  是一致支持函数;

$$(2) \forall A, B \subseteq \Theta, \text{Bel}(A \cap B) = \min(\text{Bel}(A), \text{Bel}(B)); \quad (6.1.1)$$

$$(3) \forall A, B \subseteq \Theta, \text{pl}(A \cup B) = \max(\text{pl}(A), \text{pl}(B)); \quad (6.1.2)$$

$$(4) \forall A \subseteq \Theta, A \neq \emptyset, \text{pl}(A) = \max_{\theta \in A} \text{pl}(\{\theta\}) \quad (6.1.3)$$

$$(5) \forall A \subseteq \Theta, A \neq \emptyset, Q(A) = \min_{\theta \in A} Q(\{\theta\}) \quad (6.1.4)$$

(6) 存在正整数  $n$ , 非空简单支持函数  $S_1, \dots, S_n$ , 使  $\text{Bel} = S_1 \oplus \dots \oplus S_n$ , 而且  $S_i$  的非  $\Theta$  的焦元包含在  $S_j$  的非  $\Theta$  的焦元中 ( $i < j$ )。

该定理的 (6) 说明, 任一一致支持函数是可分离支持函数, 而且也说明, 产生该支持函数的证据是无冲突的和一致的, 即尽管这些证据在改变焦元的精确度上表现出一些不同类特性, 但是它们却都针对着一个相同的目标。

另外, 一致支持函数还可以直接定义, 而不去考察它的焦元。

**定理6.1.2** 设  $\Theta$  是有限集,  $S: 2^\Theta \rightarrow [0,1]$  定义在  $\Theta$  上, 则  $S$  是一致支持函数的充分必要条件是

$$(1) S(\emptyset) = 0;$$

$$(2) S(\Theta) = 1;$$

$$(3) \forall A, B \subseteq \Theta, S(A \cap B) = \min(S(A), S(B)).$$

**定理6.1.3** 设  $\Theta$  是一个有限集, 那么一个函数  $\text{Bel}: 2^\Theta \rightarrow [0,1]$  是一致支持函数的充分必要条件是  $\text{Bel}(\Theta) = 1$ , 而且



由  $\forall A \subset \Theta, \text{Dou}(A) = \text{Bel}(\bar{A})$  所定义的函数  $\text{Dou}: 2^\Theta \rightarrow [0, 1]$  满足

$$(1) \forall A \subset \Theta, A \neq \phi, \text{Dou}(A) = \min_{\theta \in A} \text{Dou}(\{\theta\});$$

$$(2) \exists \theta \in \Theta \text{ 使得 } \text{Dou}(\{\theta\}) = 0.$$

**定理6.1.4** 设  $S_1, \dots, S_n$  是非空简单支持函数,  $A_1, \dots, A_n$  分别是其非  $\Theta$  的焦元。又设  $S = S_1 \oplus \dots \oplus S_n$  是一致的。令  $\mathcal{C}$  为  $S$  的核心, 那么集合  $A_i \cap \mathcal{C} (i=1, \dots, n)$  是套形的, 即它们可以被排列使得其中一个包含于下一个中。

设  $\text{Bel}$  是  $\Theta$  上的一致支持函数, 则由 (6.1.1) 式知

$$0 = \text{Bel}(\phi) = \text{Bel}(A \cap \bar{A}) = \min(\text{Bel}(A), \text{Bel}(\bar{A}))$$

即  $\forall A \subset \Theta$ , 或者  $\text{Bel}(A) = 0$ , 或者  $\text{Bel}(\bar{A}) = 0$ 。这说明一致支持函数不再将两个正的信度授予两个完全对立的命题。

尽管一些非一致的信度函数有时也具有这个特性, 但是在所有条件下, 只有一致支持函数才能保持这个特性。正如下定理所说。

**定理6.1.5** 一个定义在框架  $\Theta$  上的信度函数  $\text{Bel}$  是一致支持函数的充分必要条件为  $\forall A \subset \Theta$  和所有满足  $\text{Bel}(\bar{B}) < 1$  的  $B \subset \Theta$

$$\min(\text{Bel}(A | B), \text{Bel}(\bar{A} | B)) = 0$$

## §2 外形函数

一致支持函数在基于无冲突证据这一点上与贝叶斯信度函数形成了强烈的对比。但是, 在一个非常重要的方面它们却是非常相似的, 即它们都可以由单点上的似真度完全决定。在这个意义下, 这两个函数都是点函数而非集函数。

**定义6.2.1** 设  $Bel: 2^\Theta \rightarrow [0,1]$  是一个一致支持函数,  $pl$  是对应的似真度函数, 如果一个点函数  $f: \Theta \rightarrow [0,1]$  定义为

$$\forall \theta \in \Theta \quad f(\theta) = pl(\{\theta\})$$

则称  $f$  为一致支持函数  $Bel$  的外形函数 (Contour function)。

给定  $Bel$  的外形函数以后, 立即可以重新得到似真度函数  $pl$  和众信度函数  $Q$ 。由于  $Q$  和  $pl$  在单点上总是相同的, 所以由 (6.1.3) 式和 (6.1.4) 式知,

$$\begin{aligned} \forall A \subset \Theta \quad A \neq \emptyset \\ pl(A) = \max_{\theta \in A} f(\theta) \end{aligned} \quad (6.2.1)$$

$$Q(A) = \min_{\theta \in A} f(\theta) \quad (6.2.2)$$

这说明一致支持函数可由它的外形函数完全表示。因此, 一致支持函数不但可以由基本可信度分配确定, 而且可由外形函数确定。也正因为如此, 实践中的支持函数常被假设成一致支持函数。

利用外形函数可以给出一致支持函数的一个解释。

设一个一致支持函数具有5个套形焦元  $A_1 \subset A_2 \subset A_3 \subset A_4 \subset A_5$ , 其中  $A_1$  是一个单点, 每个焦元具有基本可信数  $1/5$ , 如下页图6.1和图6.2所示。

这样一个套形焦元的图形使我们想到了一个山丘, 其中  $A_1$  是山顶, 其他焦元组成了山顶周围的山坡。某一点的外形函数的值 (如图6.2所示) 对应于该点的高度, 而各焦元的边线构成了若干条等高线。但是, 这样构成的山坡不是平滑的, 而是阶梯形的, 每阶之间具有0.2的跳跃, 如:  $A_1$  高出  $A_2$  而0.2等。

这个几何解释可以帮助我们想象半流动的信质的移动情况。设我们的信质原来被局限在一个单位长度的垂直的柱

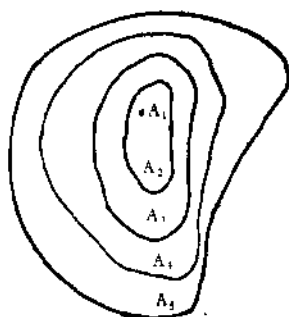


图 6.1

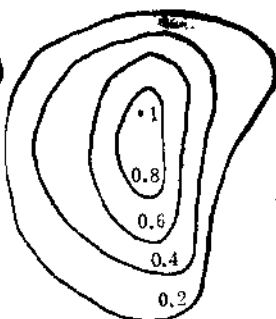


图 6.2

上，该柱立在山顶 $A_1$ 下。如果我们要求每一信质的移动局限於它原来的水平上，那么，我们就可以得到一种对信质在 $\Theta$ 上如何移动的精确表示。

### §3 证据的一致性

由于一致支持函数的简洁性，因而很多人希望在证据处理过程中使用一致支持函数。更进一步，许多人甚至认为基于证据的支持度应该总是一致的，不一致是估价证据时的错误造成的，而且这种错误也总是可以被立即纠正的。持这种观点的代表人物是著名的哲学家L.J.Cohen和经济学家G.L.S.Shackle。

在Cohen的著作中，对证据的理解似乎总是限于一致证据。这当然不能说，Cohen就拒绝不一致证据。但是，在他的整个研究过程中，证据总是针对一些可以避免证据冲突的可能性或假设。如果某个假设可以带来证据间的冲突的话，那么他就小心翼翼地避开这个假设。“当结果不能与原来的

假设相一致时，就需要改换原来的假设，以便消除这种不一致性。”\*由此可见他的“一致性”观点。

Shackle 很少谈证据，通常，他总是直接考虑我们的部分信度(partial belief)应该怎样表现。对一个特定的命题，他采用被称之为“潜在惊讶的程度”(degree of potential surprise)的量与之相联系(这个量对应于信度理论中的怀疑度 $\text{Dou}(A) = \text{Bel}(\bar{A})$ )。并且指出，潜在惊讶的程度满足定理 6.1.3 中的规则。而且他也意识到了这些规则拒绝给一命题及该命题的非同时分配正的潜在惊讶的程度。请看他说过的这段话：“将正的潜在惊讶的程度同时分配给一个命题以及它的非是一个人内心混乱的结果。”\*\*

以上这些学者都禁止在证据的估价以及信度的分配上出现冲突。换句话说，他们都拒绝处理矛盾证据。但是 Shafer 明确指出：不能期望我们的证据总是一致的。要想利用全部证据就不得不处理矛盾证据。矛盾证据在实践中是普遍存在的。<sup>[1]</sup>

不过，也有许多证据可以作为一致证据来处理，比如推理证据(inferential evidence)就可以很好地作为一致证据来处理(推理证据见下节)。

另外，由于一致证据处理起来比较方便，因而一些冲突不太厉害的证据也可以近似地作为一致证据来处理。

总之，证据间存在矛盾是普遍的，所以必须随时处理矛盾证据。但由于一致证据的简单性，因而许多证据都被作为一致证据来近似处理。

---

\* 转引自[1]

\*\* 转引自[1]

## §4 推理证据

将所有证据都作为一致证据来处理是错误的。但是，一些特殊的证据用一致证据来处理通常还是可以的。比如说，推理证据(inferential evidence)用一致证据来处理就是比较合适的。

设有一个识别框架 $\Theta$ ， $\Theta$ 中的各可能性被解释成产生某个结果的可能原因。有一批证据针对该识别框架即这批证据是用来证实该识别框架中的各可能性哪个有可能成为产生某结果的原因。那么，这批证据就称为推理证据。直观上，推理证据指向 $\Theta$ 中我们认为有很强的倾向产生结果的那些元素。这一点表明， $\Theta$ 中具有最大的这种倾向性的元素似真性更强，而具有最小的这种倾向性的元素其怀疑性更强。这种总体上的直观影响似乎用一个一致支持函数的套形焦元来描述比较好。

事实上，设 $\Theta = \{\theta_1, \dots, \theta_n\}$ ，所有的 $\theta_i$ 被排列成 $\theta_i$ 比 $\theta_{i+1}$ 具有更大的倾向产生结果( $i = 1, \dots, n-1$ )。那么 $\theta_1$ 对于产生结果来说具有最大的倾向性，因此将获得一些正的支持，从而 $\{\theta_1\}$ 将成为一个焦元。因为在产生结果的倾向性上 $\theta_2$ 比不上 $\theta_1$ ，所以 $\theta_2$ 不应该获得正的支持。但是，集合 $\{\theta_1, \theta_2\}$ 与单点集 $\{\theta_1\}$ 相比似乎应该获得更大的支持。特别是，在产生结果的倾向性上，当 $\theta_2$ 与 $\theta_1$ 非常接近而 $\theta_2$ 与 $\theta_3$ 又相差很多时，这个结论似乎更加可信。因此， $\{\theta_1, \theta_2\}$ 也将是一个焦元。如此推下去，我们将得到一个焦元的套形序列

$$\{\theta_1\} \subset \{\theta_1, \theta_2\} \subset \{\theta_1, \theta_2, \theta_3\} \subset \dots \subset \{\theta_1, \dots, \theta_{n-1}\} \subset \Theta$$

换句话说，我们得到了一个一致支持函数，该一致支持函数的外形函数 $f: \Theta \rightarrow [0, 1]$ 满足 $f(\theta_i) > f(\theta_{i+1}) (i = 1, \dots, n-1)$ 。

值得说明的是，以上讨论假定了任何两个 $\theta_i (i = 1, \dots, n-1)$ 产生结果的倾向性都严格不同。如果 $\theta_i$ 和 $\theta_{i+1}$ 具有相同的这种倾向性，那么我们将假设 $f(\theta_i) = f(\theta_{i+1})$ ，同时消去焦元 $\{\theta_i, \dots, \theta_{i+1}\}$ 。

如何估价一条推理证据呢？首要的一点就是需要考察大量的背景信息。要想找出导致一个结果的所有可能的原因，同时估价产生该结果的每一原因的倾向性，理论上需要大量的证据，实践中，又需作出大量的假设。

这里假设的作用是不容忽视的。要想不作出假设就来确定一个识别框架是不可能的。事实上，一个识别框架就是基于某一个假设而作出的。从这个意义上讲，识别框架是主观的。而这一点我们在第二章就已作了强调说明。

推理证据实际上是相当普遍的。比如说，我们经常喜欢使用推理证据来描述疑难推理。因此，在许多情形下，如果我们需要的話，即使不明显能解释成推理证据的证据，也可以解释为推理证据。

## §5 一致支持函数与可能性分布

Shafer的信度理论与L. A. Zadeh的可能性理论都是研究不确定性的理论，那么在它们之间存在着什么关系呢？本节就来探讨这种关系。

**定义6.5.1** 设论域 $X$ 是一个有限集， $\mathcal{A}$ 是 $X$ 上的一个代数， $g: \mathcal{A} \rightarrow [0, 1]$ 是定义在 $\mathcal{A}$ 上的集函数。若 $g$ 满足

$$(1) \quad g(\phi) = 0;$$

$$(2) \quad g(X) = 1;$$

$$(3) \quad \forall A \in \mathcal{A}, \forall B \in \mathcal{A}, \text{如果 } A \subset B, \text{那么 } g(A) \leq g(B).$$

则称 $g$ 为 $X$ 上的一个模糊测度。<sup>[12]</sup>

如果一个函数满足定义6.5.1的(3)则称它满足单调性。由定义6.5.1看出,模糊测度是满足单调性的集函数。

由第二章信度函数的定义可知,一个识别框架 $\Theta$ 上的信度函数 $\text{Bel}$ 满足单调性条件。即 $\forall A, B \subset \Theta$ 只要 $A \subset B$ 则 $\text{Bel}(A) \leq \text{Bel}(B)$

$$\begin{aligned} (\because \text{Bel}(B) &= \text{Bel}(A \cup (B - A)) \\ &\geq \text{Bel}(A) + \text{Bel}(B - A) \\ &\quad - \text{Bel}(A \cap (B - A)) \\ &= \text{Bel}(A) + \text{Bel}(B - A) \end{aligned}$$

$$\therefore \text{Bel}(A) \leq \text{Bel}(B)$$

所以说,信度函数 $\text{Bel}$ 是框架 $\Theta$ 上的模糊测度。这就是近年来许多学者将信度理论归于模糊数学的原因。

但是,信度函数是一个模糊测度,只是从数学形式的角度来考虑的,没有涉及任何含义问题。如果从含义上看,信度函数根本不能看成是一个模糊测度;另外,框架 $\Theta$ 与论域 $U$ 具有完全不同的含义, $\Theta$ 有时不能看成是一个纯粹数学上的集合(否则怎么会有粗化和细分,数学上的点是不能再分的!)而论域 $U$ 可以看成是一个纯数学的集合;其次, $\text{Bel}$ 的合成采用Dempster合成法则,而模糊测度却用其它的规则合成。因此,信度函数与模糊测度在本质上是有很区别的。

设 $U$ 是一个论域, $x$ 是 $U$ 上的一个变量,则对 $\forall u \in U, x$ 都可

取 $u$ 作为其值。现在, 假设 $F$ 是 $U$ 上的一个正规模糊子集, 隶属函数为 $\mu_F: U \rightarrow [0, 1]$ , 则 $F$ 可以看成是 $x$ 在 $U$ 上取值的一个模糊约束(Fuzzy restriction), 即 $x$ 只能以 $\mu_F(u)$ 的程度取到 $u$ , 换句话说 $x$ 取 $u$ 的可能性从1降到了 $\mu_F(u)$ 。从这个意义上讲,  $\mu_F$ 可称为 $x$ 取值的一个可能性分布(possibility distribution)<sup>[17]</sup>。今后可能性分布我们用 $\pi$ 表示,  $x$ 取 $A \subset U$ 的可能性用 $\Pi(A)$ 来表示(称为可能性测度), 所以

$$\forall A \subset U \quad \Pi(A) = \sup_{u \in A} \pi(u) \quad (6.5.1)$$

设 $U$ 是有限集 $U = \{x_1, \dots, x_n\}$ , 则 $U$ 上的一个可能性分布就是 $\pi: U \rightarrow [0, 1]$ , 而(6.5.1)式将退化为

$$\forall A \subset U \quad \Pi(A) = \max_{x \in A} \pi(x) \quad (6.5.2)$$

对比(6.2.1)式与(6.5.2)式可以看出,  $pl(A)$ 是以 $f(\theta)$ 为可能性分布的可能性测度, 而 $\Pi(A)$ 是以 $\pi(x)$ 为外形函数的似真度函数。

因此, 撇开意义不谈, 一致支持函数的似真度函数与可能性测度是等价的。由于一致支持函数是一般信度函数的特例, 所以可能性测度是似真度函数的特殊情况。与以上讨论相同, 尽管在数学形式上, 可能性分布等价于一致支持函数的似真度函数, 但是由于在实践意义和解释上不同, 以及采用的合成规则不同, 所以不能认为可能性理论是信度理论的特殊情况。

总之, 信度理论与模糊理论在数学形式上具有一些共同点和相似点, 但是据此决不能将两种理论混同起来, 也不能将一种理论说成是另一种理论的子集, 从而予以取消。两种理论在科学上谁也不能代替谁, 都有其特色, 都有其存在和



应用的价值。

关于这两种理论的区别与联系的研究目前仍在继续。可参看参考文献[6], [12], [13], [16]等。

## 第七章 统计证据

统计证据是证据理论在统计问题中的应用，是Shafer对统计问题的一种新的处理，是用非统计的方法研究统计问题的一种尝试。

### §1 引言

设有一个随机试验，该试验的所有可能的结果的集合为 $\mathcal{X}$ 。已知确定该试验的几率密度在类 $\{q_\theta\}_{\theta \in \Theta}$ 中， $\Theta$ 为参数空间（在这里假定 $\Theta$ 为有限集，并取为我们的识别框架）。问题是：如何找出确定该随机试验的那个几率密度，即如何确定参数 $\theta$ 的真值。

这个问题就是传统的参数估计问题。对这个问题的传统处理是：根据几次试验的观察值构造出一个统计量作为 $\theta$ 的估计值（有点估计和区间估计）。所用的方法有矩法，最大似然法，贝叶斯法，置信区间法等。这些方法的一个共同点是都要求求出一个确定的 $\theta$ 值（点值或区间值）。

对于这个问题，Shafer将试验所出现的结果 $x \in \mathcal{X}$ 看成是确定 $\theta$ 的真值的一个证据，从而求得一个 $\Theta$ 上的支持函数 $S_x$ 使

$s_x(A)$ 是观察结果 $x$ 为命题“ $\theta$ 的真值在 $\Theta$ 的子集 $A$ 中”所提供的支持度。因此,在Shafer的方法中并不要求真正求出一个 $\theta$ 的确切值,而是要求求出 $\theta$ 的确切值在某一个范围内的可能性。

## §2 单个观察值下的统计证据

Shafer的整个方法基于下面的两个约定(Convention)。

第一个约定是观察结果 $x$ 的出现应该决定一个满足 $pl(\{\theta\}) = Cq_\theta(x)$ 的似真函数(其中 $C$ 是与 $\theta$ 无关的常数)。该约定来自于这样一个直观感觉:既然 $x$ 已经出现,那么它出现的几率就应该比较大(一般认为小概率事件在一次试验中不可能发生),即观察值 $x$ 作为证据应该有利于 $\Theta$ 的把较大的几率分配给 $x$ 的那些元素,说得更详细一点,只要 $q_{\theta_0}(x) > q_{\theta_1}(x)$ 那么 $x$ 使得 $\theta \in \Theta$ 比 $\theta' \in \Theta$ 有更大的似真性(言外之意: $\theta$ 的似真性应该与 $q_\theta(x)$ 成正比)。事实上,这样一个约定也就是最大似然法所基于的最大似然原理。在最大似然法看来使 $q_\theta(x)$ 最大的 $\theta$ 就是参数 $\theta$ 的真值的估计。

第二个约定是一致性假设,即证据 $x$ 应该产生一个一致支持函数。这个约定来自于下列的直观要求:如果 $\Theta$ 的一个集合包括一些使 $q_\theta(x)$ 较小的元素 $\theta$ ,那么证据 $x$ 就不能给该子集一个比较大的支持,换句话说:为了获得一个对给定子集的一定水平的支持,那么该子集必须包含所有元素 $\theta$ ,这些元素 $q_\theta(x)$ 都超过了一定的水平。

在这两个约定下,由证据 $x$ 所决定的似真度函数的外部函数 $f: \Theta \rightarrow [0, 1]$ 为 $f(\theta) = Cq_\theta(x)$ 。因此,  $\forall A \subset \Theta$

$$pl_x(A) = \max_{\theta \in A} \{Cq_\theta(x)\} = C \max_{\theta \in A} q_\theta(x)$$

当  $A = \Theta$  时,  $pl_x(\Theta) = C \max_{\theta \in \Theta} q_\theta(x) = 1$

由此确定出

$$C = \frac{1}{\max_{\theta \in \Theta} q_\theta(x)}$$

将常数  $C$  带回上面的式子得出唯一的由  $x$  决定的似真度函数

$$\forall A \subset \Theta \quad pl_x(A) = \begin{cases} \frac{\max_{\theta \in A} q_\theta(x)}{\max_{\theta \in \Theta} q_\theta(x)} & A \neq \phi \\ 0 & A = \phi \end{cases}$$

由此我们得到一个由

$$\forall A \subset \Theta \quad S_x(A) = \begin{cases} 1 - \frac{\max_{\theta \in A} q_\theta(x)}{\max_{\theta \in \Theta} q_\theta(x)} & A \neq \phi \\ 0 & A = \phi \end{cases} \quad (7.2.1)$$

所确定的支持函数  $S_x: 2^\Theta \rightarrow [0, 1]$ 。该支持函数是证据  $x$  在框架  $\Theta$  上产生的支持函数。

### §3 统计证据的权重

上一节我们将观察值  $x$  看成是针对框架  $\Theta$  的证据, 从而得到了基于观察值  $x$  的一个一致支持函数  $S_x$ 。一致支持函数是可分离支持函数的特例, 因此对一致支持函数  $S_x$  我们可以求得它所对应的证据的估价。

**定理 7.3.1** 设一个随机试验的所有可能的结果用集合

$\mathcal{X}$  来表示, 确定该试验的几率密度在类  $\{q_\theta\}_{\theta \in \Theta}$  中,  $x \in \mathcal{X}$  为一个观察值 (而且假定至少存在一个元素  $\theta \in \Theta$  满足  $q_\theta(x) > 0$ )。又令  $\Theta = \{\theta_1, \dots, \theta_n\}$  满足  $q_{\theta_1}(x) \geq q_{\theta_2}(x) \geq \dots \geq q_{\theta_n}(x)$ , 则将观察值  $x$  作为证据所得到的框架  $\Theta$  上的一致支持函数  $S_x$  所对应的证据的估价满足下列条件

$$(1) W_x(\{\theta_1, \dots, \theta_i\}) = \log \frac{q_{\theta_1}(x)}{q_{\theta_{i+1}}(x)} \quad (i=1, \dots, n-1) \quad (7.3.1)$$

$$(2) W_x(A) = 0 \quad (A \text{ 为其他 } \Theta \text{ 的真子集})$$

$$(3) W_x(\Theta) = \begin{cases} \infty & \text{若 (7.3.1) 式对所有 } i \leq n-1 \text{ 都有限} \\ 0 & \text{其他} \end{cases}$$

(如果  $q_{\theta_1}(x) = q_{\theta_{i+1}}(x) = 0$ , 那么 (7.3.1) 式中的对数取为 0。)

由于可分离支持函数与它所对应的证据的估价是一一对应的, 因此满足定理 7.3.1 的 (1)、(2)、(3) 的估价  $W_x$  所决定的支持函数  $S_x$  必然与 (7.2.1) 式定义的  $S_x$  相同, 即估价  $W_x$  所决定的支持函数满足

$$\forall A \subset \Theta \quad A \neq \phi \quad S_x(A) = 1 - \frac{\max_{\theta \in A} q_\theta(x)}{\max_{\theta \in \Theta} q_\theta(x)}$$

所以证据  $x$  可用 (7.2.1) 式的一致支持函数来表示, 也可用定理 7.3.1 的 (1)、(2)、(3) 所定义的估价  $W_x$  来表示。

**定义 7.3.1** 设  $W: 2^\Theta \rightarrow [0, 1]$  是一个证据的估价,  $\theta, \theta' \in \Theta$ , 那么  $\sum_{\substack{\theta \in A \\ \theta' \notin A}} W(A)$  称为有利于  $\theta$  并且不利于  $\theta'$  的证据的权重。

直观上,只要 $q_{\theta}(x) > q_{\theta'}(x)$ ,那么证据 $x$ 就有利于 $\theta$ 而不利  
于 $\theta'$ 。在这种情形下,有利于 $\theta$ 而不利 $\theta'$ 的证据的权重就  
可以用 $\log \frac{q_{\theta}(x)}{q_{\theta'}(x)}$ 来表示。

因此,对于上面定义的估价 $W_x$ 要求它满足条件

$$\sum_{\substack{\theta \in A \\ \theta' \notin A}} W_x(A) = \begin{cases} \log \frac{q_{\theta}(x)}{q_{\theta'}(x)} & \text{如果 } q_{\theta}(x) > q_{\theta'}(x) \\ 0 & \text{其他} \end{cases} \quad (7.3.2)$$

就是非常自然的。事实上,我们可以证明 $W_x$ 不仅满足这个条  
件,而且该条件可以完全决定 $W_x$ 。

**定理 7.3.2** 定理 7.3.1 中定义的估价 $W_x$ 满足 (7.3.2)  
式而且是框架 $\Theta$ 上满足 (7.3.2) 式的唯一的一个估价。

所以,由 (7.2.1) 式表示的证据 $x$ 产生的一致支持函数  
 $S_x$ 所对应的证据的估价既可以由定理 7.3.1 来定义也可以由  
(7.3.2) 式来定义。

如果 $\Theta$ 只有两个元素 $\theta$ 和 $\theta'$ ,并且满足 $q_{\theta}(x) \geq q_{\theta'}(x)$ ,那  
么 (7.3.1) 式将退化为

$$W_x(\{\theta\}) = \sum_{\substack{\theta \in A \\ \theta' \notin A}} W_x(A) = \log \frac{q_{\theta}(x)}{q_{\theta'}(x)}$$

$$W_x(\{\theta'\}) = \sum_{\substack{\theta' \in A \\ \theta \notin A}} W_x(A) = 0$$

如果对于一个框架 $\Theta$ 上的估价 $W$ 来讲,除 $W(A)$  ( $A \subset \Theta$ )  
以外,其他的 $\Theta$ 的真子集上的值均为 0,那么 $W$ 将决定一个  
简单支持函数。此时 $W(A)$ 就是该简单支持函数对应的权重。  
相反,如果一个针对集 $A$ 的简单支持函数对应的证据的权重

为 $w$ ,那么该简单支持函数所对应的证据的估价 $W$ 除 $W(A) = w$ 以外,其余 $\Theta$ 的真子集上的估价值全为0。

因此,当 $\Theta$ 只有两个元素 $\theta$ 和 $\theta'$ 时,估价 $W_x$ 所决定的支持函数将退化为针对 $\{\theta\}$ ,权重为 $\log \frac{q_\theta(x)}{q_{\theta'}(x)}$ 的简单支持函数。

#### §4 多个观察值下的统计证据

当我们得到的观察值不是一个而是多个时如何处理呢?

如果所得到的观察值是 $x_1, x_2, \dots, x_n$ 的话( $x_1, x_2, \dots, x_n$ 独立),那么我们可以根据每一个 $x_i$ 求得一个 $S_{x_i}$ ,然后用 Dempster 合成法则求它们的直和 $S_{x_1} \oplus S_{x_2} \oplus \dots \oplus S_{x_n}$ ,作为基于证据 $x_1, x_2, \dots, x_n$ 的支持函数。不过该支持函数将不再是一致的。

当然得到 $x_1, x_2, \dots, x_n$ 以后,我们也可以将 $(x_1, x_2, \dots, x_n)$ 看成一个观察值(由乘积法则可知决定 $(x_1, x_2, \dots, x_n)$ 的几率密度将是 $\{q_\theta^*\}_{\theta \in \Theta}$ ,其中 $q_\theta^* = q_\theta(x_1)q_\theta(x_2) \dots q_\theta(x_n)$ )。然后可以根据上面介绍的方法直接求得一个支持函数 $S_{(x_1, x_2, \dots, x_n)}$ 作为基于证据 $x_1, x_2, \dots, x_n$ 的支持函数。显然,该支持函数仍然是一致的。

在[1]中第十一章,Shafer 比较了这两种方法的异同。特别指出了在观察值之间的冲突的处理上,这两种方法是根本对立的。第一种方法保持了关于这种冲突的信息。第二种方法却把最后的支持函数强加上一致性,因此隐藏了任何冲突的信息。换句话说第一种方法比第二种方法更适合于表示证

据的冲突。

## §5 统计证据的折扣

前几章我们总是假设包含在证据中的不确定性或者由框架 $\Theta$ 的假设所承担,或者通过支持函数 $S: 2^\Theta \rightarrow [0, 1]$ 来说明。但是设想下列一种情况是可能的,即因为一个支持函数不能说明反映整个证据的某些特殊的不确定性,所以,一个支持函数 $S: 2^\Theta \rightarrow [0, 1]$ 似乎是不精确的。因此,在这种情况下折扣由 $S$ 给定的支持度将是自然的。事实上,如果我们对整个证据只有 $1-\alpha$ 的似真度,其中 $0 \leq \alpha \leq 1$ ,那么我们可以采用 $\alpha$ 作为我们的折扣率,对 $\Theta$ 的每一个真子集的支持度从 $S(A)$ 减少到 $(1-\alpha)S(A)$ 。

**定理7.4.1** 设 $\text{Bel}: 2^\Theta \rightarrow [0, 1]$ 是一个信度函数,  $0 < \alpha < 1$ 。又设 $\text{Bel}^\alpha: 2^\Theta \rightarrow [0, 1]$ 满足

$$\text{Bel}^\alpha(\Theta) = 1$$

$$\forall A \subset \Theta \quad A \neq \Theta \quad \text{Bel}^\alpha(A) = (1-\alpha)\text{Bel}(A)$$

则 $\text{Bel}^\alpha$ 是一个信度函数。

设折扣率为 $\alpha$ ,那么证据对 $\Theta$ 的每一个真子集的支持度全部由 $S(A)$ 减少到 $(1-\alpha)S(A)$ 。因此每一个真子集的支持度的折扣都是相同的。那么对于焦元的基本可信度分配,其折扣又是什么情况呢?

**定理7.4.2** 设 $\text{Bel}^\alpha: 2^\Theta \rightarrow [0, 1]$ 是由 $\text{Bel}: 2^\Theta \rightarrow [0, 1]$ 折扣 $\alpha$ 所得来的(即 $\text{Bel}^\alpha$ 按照定理7.4.1定义),那么 $\text{Bel}^\alpha$ 对应的基本可信度分配 $m^\alpha: 2^\Theta \rightarrow [0, 1]$ 满足

$$m^\alpha(\Theta) = (1-\alpha)m(\Theta) + \alpha$$



$$\forall A \subset \Theta \quad A \neq \Theta \quad m^\alpha(A) = (1 - \alpha)m(A)$$

该定理说明,除 $\Theta$ 以外,其余焦元的基本可信度分配也都折扣 $1 - \alpha$ 。

以上所讨论的折扣的概念对任何信度函数或支持函数都是适用的。但是在表示统计证据的不确定性( $x$ 的出现是不确定的)时能起到更好的作用。

当 $x$ 的出现依赖于一种经验的时候,那么我们可以认为我们观察到了 $x$ ,但是是不一定的。在这种情况下,折扣由 $x$ 的出现所得到的 $S_x: 2^\Theta \rightarrow [0, 1]$ 将是非常自然的。

假设得到的观察结果为 $x_1, x_2, \dots, x_n$ ,由此所得到的支持函数为 $S_{x_1}, S_{x_2}, \dots, S_{x_n}$ 。又设 $S_{x_1}$ 与 $S_{x_2}, \dots, S_{x_n}$ 具有强烈的冲突,而 $S_{x_2}, \dots, S_{x_n}$ 之间不存在强的冲突,那么 $x_1$ 被认为是一个不一致观察(discrepant observation)或一个边远观察(outlier)。这种不一致观察的影响用折扣来描述似乎也是比较好的。

## §6 在框架的相容变换下 支持函数的变化

以上我们所讨论的统计证据,即从一个给定的随机试验中所得到的观察结果,仅仅是相对于一个特定的框架 $\Theta$ 而言的。但是也存在着这样的情况,该统计证据可以被多个相容框架所识别。那么在这种情形下,该统计证据在这多个相容框架上所产生的支持函数有什么关系呢?

虽然我们有时说统计规范(statistical specification) $\{q_\theta\}_{\theta \in \Theta}$ 和框架 $\Theta$ 相联系,但是 $\{q_\theta\}_{\theta \in \Theta}$ 相对于 $\Theta$ 而言是固有

的, 内在的, 即  $\{q_\theta\}_{\theta \in \Theta}$  并不随着框架  $\Theta$  的变化而变化。事实上, 确定一个随机试验的几率密度应该是唯一的。由此, 我们可以得到下面的直观假定。

设  $\omega: 2^{\mathcal{X}} \rightarrow 2^{\mathcal{X}}$  是一个细分,  $\theta$  是  $\Theta$  的一个元素,  $q: \mathcal{X} \rightarrow [0, 1]$  是决定一个给定的随机试验的几率密度, 那么下面的两种说法等价

(1) 如果与  $\Theta$  相联系的参数的真值是  $\theta$ , 那么  $q$  决定该试验;

(2) 如果与  $\Theta$  相联系的参数的真值在  $\omega(\{\theta\})$  中, 那么  $q$  决定该试验。

根据该假定, 如果我们知道了与一个框架相联系的某试验的规范。那么我们不仅可以说出与其他相容框架相联系的规范是否存在, 而且如果存在的话也能知道它是什么。请看下面的定理。

**定理 7.6.1** 假设统计规范  $\{q_\theta\}_{\theta \in \Theta}$  与框架  $\Theta$  相联系,  $\Theta'$  是框架  $\Theta$  的一个相容框架。那么存在与  $\Theta'$  相联系的同一试验的统计规范的充分必要条件为: 只要存在  $\Theta'$  的一个元素与  $\theta_1$  和  $\theta_2$  都相容, 那么  $q_{\theta_1} = q_{\theta_2}$  (即只要存在  $\theta' \in \Theta'$  满足对应于  $\{\theta'\}$  的命题与对应于  $\{\theta_1\}$  的命题和对应于  $\{\theta_2\}$  的命题都相容, 则  $q_{\theta_1} = q_{\theta_2}$ )。

该定理的主要作用由下面的两个简单推论来体现。

**推论 1** 假设  $\{q_\theta\}_{\theta \in \Theta}$  与  $\Theta$  相联系且  $\omega: 2^{\mathcal{X}} \rightarrow 2^{\mathcal{X}}$  是一个细分, 那么存在一个与  $\Theta$  相联系的规范, 并且, 事实上,  $q_\theta$  与  $\omega(\{\theta\})$  的每一个元素相联系。

**推论 2** 假设  $\{q_\theta\}_{\theta \in \Theta}$  与  $\Theta$  相联系, 而且  $\omega: 2^{\mathcal{X}} \rightarrow 2^{\mathcal{X}}$  是一个细分, 那么如果存在  $\theta_0 \in \Theta_0$  并且  $\theta, \theta' \in \omega(\{\theta_0\})$  满足

$q_0 \neq q_1$ , 则不存在与  $\Theta_0$  相联系的规范。

下面的两个定理回答了本节开头提出的问题, 即统计证据在相容框架上产生的支持函数具有什么样的关系。

**定理 7.6.2** 假设  $\omega: 2^\Theta \rightarrow 2^\Omega$  是一个细分,  $\mathcal{X}$  是一个给定的随机试验的可能结果集。又设存在一个与  $\Theta$  相联系的试验的统计规范 (因而由定理 7.6.1 知也存在与  $\Omega$  相联系的统计规范)。令  $S_x^\Theta$  与  $S_x^\Omega$  分别表示根据 §2 的方法由结果  $x \in \mathcal{X}$  决定的  $\Theta$  与  $\Omega$  上的支持函数, 那么  $S_x^\Omega$  是  $S_x^\Theta$  的空扩张。

**定理 7.6.3** 设  $\Theta_1$  和  $\Theta_2$  是相容框架,  $\mathcal{X}$  是一个给定的随机试验的可能结果集。又设每一个框架具有一个与它相联系的试验的统计规范。令  $S_x^{\Theta_1}$  和  $S_x^{\Theta_2}$  分别表示根据 §2 的方法由  $x \in \mathcal{X}$  决定的  $\Theta_1$  与  $\Theta_2$  上的支持函数, 那么  $S_x^{\Theta_1}$  与  $S_x^{\Theta_2}$  相容。

在 [1] 中, Shafer 还讨论了统计证据的理论与贝叶斯统计理论的比较问题, 以及假设对于框架  $\Theta$  的作用问题等, 限于篇幅, 不再详述。

## 第八章 理论上的进一步探讨

这一章是作者在理论上所作的探讨。

### §1 证据合成的公理化研究

本节首先将代数系统中运算的概念从映射推广到0—1对应<sup>[10]</sup>，得到了定义不完全的运算的概念。由此又得到了定义不完全的代数系统的概念。然后讨论了定义不完全的代数系统所具有的一些性质以及具备某些性质的一些特殊的定义不完全的代数系统，诸如：定义不完全的半群、群、环、域等的概念。最后，由这种定义不完全的代数系统的概念引出了研究证据合成的公理化方法。

#### 一、引言

在代数系统的定义中，除要求一个非空集合  $S$  以外，还要求一个定义在此集合上的运算。所谓一个  $n$  元运算就是一个从  $S^n$  到  $S$  的映射  $f: S^n \rightarrow S$ ，所以对于  $S$  中的任意  $n$  个元素都存在  $S$  中的一个唯一的元素与之对应。换句话说， $S$  中任意  $n$  个元素都要求能进行运算。但是在实际问题中，有时却要求讨论，在一个非空集合上所定义的运算对一些元素有意

义,对另外一些元素无意义的情况,此时就要用到定义不完全的运算及定义不完全的代数系统的概念。

## 二、概念和性质

引入一个记号:  $A \Longleftrightarrow B$  表示:  $A$  存在的充分必要条件是  $B$  存在,且当它们存在时  $A=B$ 。

**定义8.1.1** 当集合  $X$  的每一个元素通过某种关系对应于集合  $Y$  中不多于一个的元素(可能没有)时,这种集合  $X$  和  $Y$  元素之间的关系称为0—1对应<sup>[10]</sup>。

**定义8.1.2** 设  $A$  为任意一个集合,则  $A^n$  到  $A$  的一个0—1对应称为  $A$  中的一个定义不完全的  $n$  元运算( $n$  为任意自然数)。

**定义8.1.3** 对任意一个非空集合  $A$ ,若在  $A$  中存在一些定义不完全的  $n$  元运算  $\oplus, \otimes, \dots, \circledast$ , 则  $(A, \oplus, \otimes, \dots, \circledast)$  称为一个定义不完全的代数系统。

显然,映射是0—1对应的特例,因此运算是定义不完全的运算的特例,而代数系统又是定义不完全的代数系统的特例。

与通常意义下的代数系统<sup>[20]</sup>一样,定义不完全的代数系统也有以下一些常见的性质。以下所讨论的运算均假设为二元运算。

**定义8.1.4 (结合律)** 一个定义不完全的代数系统  $(A, \oplus)$ , 若  $\forall x, y, z \in A$

$$(x \oplus y) \oplus z \Longleftrightarrow x \oplus (y \oplus z)$$

则称该运算是可结合的。

**定义8.1.5 (交换律)** 设  $(A, \oplus)$  为一个定义不完全的代数系统, 若  $\forall x, y \in A$

$$x \oplus y \iff y \oplus x$$

则称该运算为可交换的。

**定义8.1.6 (分配律)** 设  $(A, \circ, \cdot)$  为一个定义不完全的代数系统。若  $\forall x, y, z \in A$

$$x \circ (y \cdot z) \iff (x \circ y) \cdot (x \circ z)$$

则称运算  $\circ$  对运算  $\cdot$  满足第一分配律。同理可定义运算  $\cdot$  对运算  $\circ$  满足第一分配律。

若  $\forall x, y, z \in A$

$$(y \cdot z) \circ x \iff (y \circ x) \cdot (z \circ x)$$

则称运算  $\circ$  对运算  $\cdot$  满足第二分配律。同理可定义运算  $\cdot$  对运算  $\circ$  满足第二分配律。

若运算  $\circ$  对运算  $\cdot$  既满足第一分配律，又满足第二分配律，则称运算  $\circ$  对运算  $\cdot$  满足分配律。

**定义8.1.7 (幺元)** 设  $(A, \oplus)$  为一个定义不完全的代数系统。若  $\exists e \in A$ ，使得

$$\forall x \in A, x \oplus e = e \oplus x = x$$

则称  $e$  为该运算的一个幺元。

**定理8.1.1** 一个定义不完全的代数系统  $(A, \oplus)$  的运算  $\oplus$  的幺元若存在，则必唯一。

证：假设存在幺元  $e_1$  和  $e_2$ ，则

$$e_1 = e_2 \oplus e_1 = e_1 \oplus e_2 = e_2 \quad \text{证毕}$$

**定义8.1.8 (零元)** 设  $(A, \circ)$  为一个定义不完全的代数系统，若存在一个元素  $0 \in A$  使得

$$\forall x \in A \quad 0 \circ x = x \circ 0 = 0$$

则称它为运算  $\circ$  的零元。

**定义8.1.9 (逆元)** 设  $(A, \circ)$  为一个定义不完全的代

数系统。若对元素  $x, \exists x_l^{-1} \in A$  使得

$$x_l^{-1} \circ x = e$$

则称  $x_l^{-1}$  为  $x$  对运算  $\circ$  的左逆元。如果对  $A$  中元素  $x, \exists x_r^{-1} \in A$  使得

$$x \circ x_r^{-1} = e$$

则称  $x_r^{-1}$  为  $x$  对运算  $\circ$  的右逆元。

**定理 8.1.2** 一个定义不完全的代数系统  $(A, \circ)$  如果其运算  $\circ$  满足结合律, 则其左右逆元相等。

证: 设  $x_l^{-1}, x_r^{-1}$  分别为  $x$  的左右逆元, 则由定义

$$x_l^{-1} \circ x = e$$

$$x \circ x_r^{-1} = e$$

$$\therefore x_l^{-1} = x_l^{-1} \circ e = x_l^{-1} \circ (x \circ x_r^{-1})$$

$$= (x_l^{-1} \circ x) \circ x_r^{-1} = e \circ x_r^{-1} = x_r^{-1} \quad \text{证毕}$$

在两个定义不完全的代数系统之间, 也可以定义同构与同态的概念。

**定义 8.1.10** 设  $(X, \circ)$  与  $(Y, \bullet)$  为两个定义不完全的代数系统,  $\circ$  和  $\bullet$  分别为它们上的定义不完全的二元运算。如果存在一个一一对应的映射  $g: X \rightarrow Y$ , 使得

(1)  $\forall x_1, x_2 \in X$ , 若  $x_1 \circ x_2$  存在, 则  $g(x_1) \bullet g(x_2)$  也存在, 且

$$g(x_1 \circ x_2) = g(x_1) \bullet g(x_2)$$

(2)  $\forall y_1, y_2 \in Y$ , 若  $y_1 \bullet y_2$  存在, 则  $g^{-1}(y_1) \circ g^{-1}(y_2)$  也存在, 且

$$g^{-1}(y_1 \bullet y_2) = g^{-1}(y_1) \circ g^{-1}(y_2)$$

则称  $g$  为从  $(X, \circ)$  到  $(Y, \bullet)$  的同构映射,  $(X, \circ)$  和  $(Y, \bullet)$  称为同构。

对两个定义不完全的同构的代数系统 $(X, \circ)$ 和 $(Y, \bullet)$ , 下列性质成立。

**定理8.1.3** 如果 $(X, \circ)$ 满足结合律, 则 $(Y, \bullet)$ 也必满足结合律。

证: 设  $y_1, y_2, y_3 \in Y$  则  $\exists x_1, x_2, x_3 \in X$  使得

$$g(x_1) = y_1 \quad g(x_2) = y_2 \quad g(x_3) = y_3$$

又设  $(y_1 \bullet y_2) \bullet y_3$  存在, 则由定义知  $g^{-1}(y_1) \circ g^{-1}(y_2)$  存在, 且

$$g^{-1}(y_1 \bullet y_2) = g^{-1}(y_1) \circ g^{-1}(y_2)$$

$$\therefore g^{-1}(y_1 \bullet y_2) \circ g^{-1}(y_3)$$

$$\text{即 } (g^{-1}(y_1) \circ g^{-1}(y_2)) \circ g^{-1}(y_3)$$

$$\text{即 } (x_1 \circ x_2) \circ x_3 \text{ 存在}$$

$$\because (X, \circ) \text{ 满足结合律} \quad \therefore x_1 \circ (x_2 \circ x_3) \text{ 存在}$$

同样, 由定义知  $g(x_2) \bullet g(x_3)$  存在且

$$g(x_2 \circ x_3) = g(x_2) \bullet g(x_3)$$

$$\therefore g(x_1) \bullet g(x_2 \circ x_3)$$

$$\text{即 } g(x_1) \bullet (g(x_2) \bullet g(x_3))$$

$$\text{即 } y_1 \bullet (y_2 \bullet y_3) \text{ 存在}$$

同理当  $y_1 \bullet (y_2 \bullet y_3)$  存在时,  $(y_1 \bullet y_2) \bullet y_3$  也存在。

又当它们都存在时,

$$(y_1 \bullet y_2) \bullet y_3 = (g(x_1) \bullet g(x_2)) \bullet g(x_3)$$

$$= g((x_1 \circ x_2) \circ x_3)$$

$$y_1 \bullet (y_2 \bullet y_3) = g(x_1) \bullet (g(x_2) \bullet g(x_3))$$

$$= g(x_1 \circ (x_2 \circ x_3))$$

$$\therefore (x_1 \circ x_2) \circ x_3 = x_1 \circ (x_2 \circ x_3)$$

$$g: X \rightarrow Y \text{ 是一一对应的}$$



$$\therefore g((x_1 \circ x_2) \circ x_3) = g(x_1 \circ (x_2 \circ x_3))$$

$$\therefore (y_1 * y_2) * y_3 = y_1 * (y_2 * y_3) \quad \text{证毕}$$

**定理8.1.4** 如果 $(X, \circ)$ 满足交换律, 则 $(Y, *)$ 也满足交换律。

证明略。

**定理8.1.5** 设 $(X, \circ)$ 存在零元 $0_x$ , 则 $(Y, *)$ 也必存在零元 $0_y$ , 且有 $g(0_x) = 0_y$ 。

证:

$\because 0_x$ 为 $(X, \circ)$ 的零元

$$\therefore \forall x \in X \text{ 有 } 0_x \circ x = x \circ 0_x = 0_x$$

$$\because \forall y \in Y \quad \exists x \in X \text{ 使 } g(x) = y$$

$$\therefore g(0_x) * y = g(0_x) * g(x) \\ = g(0_x \circ x) = g(0_x)$$

$$y * g(0_x) = g(x) * g(0_x) \\ = g(x \circ 0_x) = g(0_x)$$

$$\therefore \forall y \in Y \quad \exists g(0_x) \in Y \text{ 使得}$$

$$g(0_x) * y = y * g(0_x) = g(0_x)$$

$$\therefore g(0_x) \text{ 为 } (Y, *) \text{ 的零元} \quad \text{证毕}$$

**定义8.1.11** 设有两个定义不完全的代数系统 $(X, \circ)$ 与 $(Y, *)$ , 其中 $\circ$ 与 $*$ 都是二元运算。如果存在一个映射 $g: X \rightarrow Y$ , 使得对任意的 $x_1, x_2 \in X$ , 若 $x_1 \circ x_2$ 存在, 则 $g(x_1) * g(x_2)$ 也存在, 而且

$$g(x_1 \circ x_2) = g(x_1) * g(x_2)$$

则称 $g$ 是从 $(X, \circ)$ 到 $(Y, *)$ 的同态映射, 而 $(X, \circ)$ 与 $(Y, *)$ 称为同态。

下面定义一些定义不完全的特定的代数系统。

**定义8.1.12** 若定义不完全的代数系统  $(A, \oplus)$  中的二元运算  $\oplus$  满足可结合性, 则称该系统为定义不完全的半群。对于一个定义不完全的半群  $(A, \oplus)$  来说, 若运算  $\oplus$  还存在幺元, 则称该系统为定义不完全的含幺半群; 若运算  $\oplus$  还满足可交换性, 则称为定义不完全的可交换半群; 若既满足可交换性又存在幺元, 则称该系统为定义不完全的可交换含幺半群。

**定义8.1.13** 一个定义不完全的代数系统  $(A, \oplus)$  ( $\oplus$  为二元运算), 如果满足下列条件

(1) 结合律: 即  $\forall x, y, z \in A$

$$(x \oplus y) \oplus z \iff x \oplus (y \oplus z)$$

(2) 存在幺元: 即  $\exists e \in A$  使得

$$\forall x \in A \quad x \oplus e = e \oplus x = x$$

(3) 几乎处处存在逆元: 即对  $A$  内除了可以和幺元  $e$  运算以外, 还至少可以和除  $e$  以外的一个其他元素运算的任意元素  $a$ , 均有一个  $a^{-1} \in A$ , 使得

$$a \oplus a^{-1} = a^{-1} \oplus a = e$$

则称此代数系统  $(A, \oplus)$  为定义不完全的群。

**定义8.1.14** 一个定义不完全的群  $(A, \oplus)$ , 如果满足交换律, 则称为定义不完全的可交换群或定义不完全的阿贝尔群。

**定义8.1.15** 一个定义不完全的代数系统  $(A, \oplus, \otimes)$ , 如果它的两个定义不完全的二元运算满足下列条件

(1)  $(A, \oplus)$  是一个定义不完全的可交换群;

(2)  $(A, \otimes)$  是一个定义不完全的半群;

(3) 运算  $\otimes$  对  $\oplus$  满足分配律, 即  $\forall x, y, z \in A$

$$x \otimes (y \oplus z) \iff (x \otimes y) \oplus (x \otimes z)$$

$$(y \oplus z) \otimes x \iff (y \otimes x) \oplus (z \otimes x)$$

则称  $(A, \oplus, \otimes)$  为定义不完全的环。

**定义 8.1.16** 一个定义不完全的环  $(F, +, \cdot)$  (其中  $F$  至少包含一个以上的元素), 如果它的两个定义不完全的二元运算  $+$  和  $\cdot$  满足下列条件

(1)  $(F, \cdot)$  有幺元;

(2)  $(F, \cdot)$  是可交换的;

(3)  $(F, \cdot)$  除零元以及只能和幺元、零元进行运算的元素以外的所有元素均有逆元素。

则称  $(F, +, \cdot)$  为定义不完全的域。

**定义 8.1.17** 定义不完全的代数系统  $(B, +, \cdot, \neg)$ , 如果定义不完全的二元运算  $+$  与  $\cdot$  及定义不完全的一元运算  $\neg$  满足下列条件

(1) 交换律: 即  $\forall x, y \in B$

$$x + y \iff y + x$$

$$x \cdot y \iff y \cdot x$$

(2) 结合律: 即  $\forall x, y, z \in B$

$$(x + y) + z \iff x + (y + z)$$

$$(x \cdot y) \cdot z \iff x \cdot (y \cdot z)$$

(3) 分配律: 即  $\forall x, y, z \in B$

$$x \cdot (y + z) \iff (x \cdot y) + (x \cdot z)$$

$$(y + z) \cdot x \iff (y \cdot x) + (z \cdot x)$$

$$x + (y \cdot z) \iff (x + y) \cdot (x + z)$$

$$(y \cdot z) + x \iff (y + x) \cdot (z + x)$$

(4) 存在幺元: 对运算  $+$  存在幺元  $0$ , 对运算  $\cdot$  存在幺

元 1。

(5) 互补性: 即  $\forall x \in B$ , 若  $\bar{x}$  及  $x + \bar{x}$  存在, 则

$$x + \bar{x} = 1$$

若  $\bar{x}$  及  $x \cdot \bar{x}$  存在, 则

$$x \cdot \bar{x} = 0$$

则称  $(B, +, \cdot, \bar{\phantom{x}})$  为定义不完全的布尔代数。

关于以上定义的这些定义不完全的代数系统的性质, 限于篇幅, 不再讨论。

### 三、研究证据合成的公理化方法

Shafer 在 [1] 中阐述了证据合成的 Dempster 法则。两个信度函数合成的 Dempster 法则可以表述为:

设  $Bel_1$  与  $Bel_2$  是基于不同证据的两个信度函数, 其焦点分别为  $A_1, \dots, A_n$  和  $B_1, \dots, B_l$ 。则当

$$\sum_{A_i \cap B_j = \emptyset} m_1(A_i) m_2(B_j) < 1 \quad (8.1.1)$$

时,  $Bel_1$  与  $Bel_2$  的合成具有如下的基本可信度分配

$$m(A) = \begin{cases} \frac{\sum_{A_i \cap B_j = A} m_1(A_i) m_2(B_j)}{1 - \sum_{A_i \cap B_j = \emptyset} m_1(A_i) m_2(B_j)} & A \neq \emptyset \\ 0 & A = \emptyset \end{cases}$$

可见, 对于任意两个满足 (8.1.1) 式的信度函数来讲, 利用 Dempster 合成法则都可以求得一个特定的信度函数作为原先那两个信度函数的合成。

但是, 证据的情况千变万化, 证据的合成也应因人而异, 所以用一种很生硬的规则去概括全人类进行证据合成时

所遵循的原则未免有点牵强附会。不过，Dempster 合成法则还是一种合成法则，只不过对某些人适用，对其他人不适用罢了。因此，作者认为研究合成最好的办法应该是通过几种特殊的合成法则的考察找出人类进行证据合成时所遵循的普遍原则，即从公理化的角度来研究证据的合成。下面我们就通过对 Dempster 合成法则的考察给出一般合成的公理化定义。

设  $\Theta$  为一识别框架，则该框架上所有的信度函数形成一集合，用  $\mathcal{B}$  来表示。设  $\oplus$  为  $\mathcal{B}$  上两元素的直和运算。可以验证，空间  $(\mathcal{B}, \oplus)$  (称为信度合成空间) 为一个定义不完全的可交换含么半群。即定义不完全的运算  $\oplus$  满足

(1) 可结合性：即  $\forall Bel_1, Bel_2, Bel_3 \in \mathcal{B}$

$$(Bel_1 \oplus Bel_2) \oplus Bel_3 \iff Bel_1 \oplus (Bel_2 \oplus Bel_3)$$

(2) 可交换性：即  $\forall Bel_1, Bel_2 \in \mathcal{B}$

$$Bel_1 \oplus Bel_2 \iff Bel_2 \oplus Bel_1$$

(3) 存在么元  $Bel_0$  ( $Bel_0$  为空信度函数，即其焦元只有  $\emptyset$ )，即  $\forall Bel \in \mathcal{B}$  都有

$$Bel \oplus Bel_0 = Bel_0 \oplus Bel = Bel$$

其中 (1) 可以用 Dempster 合成法则的众信度表示法得到证明；(2) 从合成后的基本可信度分配的表达式中一看即明；(3) 显然成立。

我们可以看到 (1)、(2)、(3) 不仅是 Dempster 合成法则所满足的条件，也是人类进行证据合成时所遵循的基本原则，可以作为定义一般合成的基本公理。

**定义 8.1.18** 设  $\mathcal{B} = \{Bel \mid Bel \text{ 为识别框架 } \Theta \text{ 上的信度函数}\}$ ， $\oplus$  为  $\mathcal{B}$  上的一个定义不完全的二元运算。 $\forall Bel_1,$

$Bel_2 \in \mathcal{B}$ , 设  $Bel_1$  与  $Bel_2$  所对应的基本可信度分配分别为  $m_1$  和  $m_2$ ,  $A_1, \dots, A_n$  和  $B_1, \dots, B_l$  分别为  $Bel_1$  与  $Bel_2$  的焦点。若

$$\sum_{A_i \cap B_j \neq \emptyset} m_1(A_i) m_2(B_j) < 1$$

则规定  $Bel_1$  与  $Bel_2$  之间可以施以运算  $\oplus$ ; 而且假定  $(\mathcal{B}, \oplus)$  形成了一个定义不完全的可交换含么半群, 即定义不完全的二元运算  $\oplus$  满足

- (1) 可结合性: 即  $\forall Bel_1, Bel_2, Bel_3 \in \mathcal{B}$   
 $(Bel_1 \oplus Bel_2) \oplus Bel_3 \iff Bel_1 \oplus (Bel_2 \oplus Bel_3)$
- (2) 可交换性: 即  $\forall Bel_1, Bel_2 \in \mathcal{B}$   
 $Bel_1 \oplus Bel_2 \iff Bel_2 \oplus Bel_1$
- (3) 空信度函数  $Bel_0$  为其么元: 即  $\forall Bel \in \mathcal{B}$   
 $Bel \oplus Bel_0 = Bel_0 \oplus Bel = Bel$

则称空间  $(\mathcal{B}, \oplus)$  为信度合成空间, 定义不完全的运算  $\oplus$  为信度的合成。

有了该定义, 以后对于两个信度函数之间的任何定义不完全的运算只要满足 (1)、(2)、(3), 则就是一种特殊的证据合成法则。

值得注意的是 (1)、(2)、(3) 对于信度的合成也许是不够的, 即由 (1)、(2)、(3) 定义的运算很可能太广。但是, 这三条对于信度的合成却是基本的。可以断言, 如果一种运算不满足这三条, 那么它肯定不是合成运算。究竟需要几条规则 (公理) 才能恰到好处地反映出人类进行证据合成的本质, 这需要进一步研究才能真正弄清。

## §2 多层证据处理

在似真推理与可信度判断中，我们（决策者）往往根据一批证据对某一命题（或某些命题）产生一个信度。但是，有时会发生这样的情况：对用来作为证据的命题，我们也不是百分之百的全信。在这种情况下，我们通常的一种做法就是再进一步寻求证据，以证明我们刚才用作证据的命题是否为真。如果现在得到的证据我们还不是百分之百的全信，那么我们就再寻求第三批证据，以此类推。这样为了证明某一命题为真我们找到多批证据，而这些证据又都处在不同的层次上，因此我们把这种证据处理，称为多层证据处理。处在各个层次上的证据依次称为第一层证据，第二层证据等。

这一问题也可陈述为反向推理的形式，因而也可称为逆推理证据。为了判断一些命题的真假，我们首先作出一些假设作为证据，在这些假设下，原命题为真的程度可以给出，然后对这些假设再寻求证据以求证明。在寻求新证据的过程中又可以作出新的假设，以此类推。这种方法也是一种已知结论（假设）找前提然后与事实相匹配的推理方法。

从上面的叙述可以看出，多层证据的处理似乎才真正符合人类处理证据的机制。

在[1]中，Shafer 为我们提供了处理单层证据的有力工具。那么对多层证据的处理又该怎么样呢？

最简单的情况就是：一批证据  $E_1$  以  $s_1$  的程度支持着命题  $A$ ，另一批证据  $E_2$  以  $s_2$  的程度支持着  $E_1$ ，则直观上可得到  $E_2$  以  $s_1 \cdot s_2$  的程度支持着  $A$ 。用 Shafer 证据处理方法可

陈述为：证据  $E_1$  针对着框架  $\Theta$  中的子集  $A$ ，因此在  $\Theta$  中产生了一个基本可信度分配  $m_1$

$$m_1(B) = \begin{cases} s_1 & B = A \\ 1 - s_1 & B = \Theta \\ 0 & \text{其他} \end{cases}$$

又证据  $E_2$  针对着框架  $\Omega$  中的  $\bar{A}$ （对应于  $E_1$ ），因而在  $\Omega$  上也产生了一个基本可信度分配  $m_2$

$$m_2(B) = \begin{cases} s_2 & B = \bar{A} \\ 1 - s_2 & B = \Omega \\ 0 & \text{其他} \end{cases}$$

所以，由于  $E_1$  的传递作用， $E_2$  对  $A$  的作用就可以用基本可信度分配  $m$  来表示

$$m(B) = \begin{cases} s_1 s_2 & B = A \\ 1 - s_1 s_2 & B = \Theta \\ 0 & \text{其他} \end{cases}$$

一般情况下，在框架  $\Theta$  上， $E_1$  针对  $A_1, \dots, A_n$  产生了一个基本可信度分配  $m_1$

$$m_1(B) = \begin{cases} a_1 & B = A_1 \\ \vdots & \vdots \\ a_n & B = A_n \\ 1 - \sum_{i=1}^n a_i & B = \Theta \\ 0 & \text{其他} \end{cases}$$

在框架  $\Omega$  上， $E_2$  针对  $B_1, \dots, B_n$  ( $B_1, \dots, B_n$  对应于  $E_1$ ，下同) 产生了一个基本可信度分配  $m_2$



$$m_2(B) = \begin{cases} b_1 & B = B_1 \\ \vdots & \vdots \\ b_n & B = B_n \\ 1 - \sum_{i=1}^n b_i & B = \Omega \\ 0 & \text{其他} \end{cases}$$

则从直观上可知, 由于  $E_1$  传递  $E_2$  在  $\Theta$  上产生的基本可信度分配为

$$m(B) = \begin{cases} a_1 b_1 & B = A_1 \\ \vdots & \vdots \\ a_n b_n & B = A_n \\ 1 - \sum_{i=1}^n a_i b_i & B = \Theta \\ 0 & \text{其他} \end{cases}$$

$$\begin{aligned} \because \sum_{i=1}^n a_i b_i &= a_1 b_1 + \dots + a_n b_n \\ &\leq \sum_{i=1}^n a_i \cdot \sum_{i=1}^n b_i \leq 1 \cdot 1 = 1 \end{aligned}$$

$$\therefore m(\Theta) = 1 - \sum_{i=1}^n a_i b_i \geq 0$$

$$\text{又} \quad \sum_{i=1}^n m(A_i) + m(\Theta) = 1$$

所以这种定义是合理的。

当  $E_2$  并不完全针对着  $E_1$ , 还针对着其他的命题时, 问题如何处理?

设  $E_1$  在  $\Theta$  上产生了与前相同的基本可信度分配  $m_1$ ,  $E_2$  在  $\Omega$  上产生的基本可信度分配  $m_2$  为

$$m_2(B) = \begin{cases} b_1 & B = B_1 \\ \vdots & \vdots \\ b_n & B = B_n \\ b_{n+1} & B = B_{n+1} \\ \vdots & \vdots \\ b_{n+i} & B = B_{n+i} \\ 1 - \sum_{i=1}^{n+i} b_i & B = \Omega \\ 0 & \text{其他} \end{cases}$$

因为  $B_{n+1}, \dots, B_{n+i}$  所对应的命题与我们所关心的问题无关, 所以不予考虑。  $E_2$  由于  $E_1$  的传递在  $\Theta$  上产生的基本可信度分配  $m$  仍为

$$m(B) = \begin{cases} a_1 b_1 & B = A_1 \\ \vdots & \vdots \\ a_n b_n & B = A_n \\ 1 - \sum_{i=1}^n a_i b_i & B = \Theta \\ 0 & \text{其他} \end{cases}$$

当  $E_2$  所针对的命题包含于  $E_1$  时, 又该怎样处理呢?

设  $E_1$  在  $\Theta$  上产生的基本可信度分配  $m_1$  与前相同,  $E_2$  在  $\Omega$  上产生的基本可信度分配  $m_2$  为

$$m_2(B) = \begin{cases} b_1 & B = B_1 \\ \vdots & \vdots \\ b_l & B = B_l (l < n) \\ 1 - \sum_{i=1}^l b_i & B = \Omega \\ 0 & \text{其他} \end{cases}$$

此时,  $E_2$  由于  $E_1$  的传递在  $\Theta$  上产生的基本可信度分配  $m$  为

$$m(B) = \begin{cases} a_1 b_1 & B = A_1 \\ \vdots & \vdots \\ a_i b_i & B = A_i \\ \vdots & \vdots \\ a_{i+1} & B = A_{i+1} \\ \vdots & \vdots \\ a_n & B = A_n \\ 1 - \sum_{i=1}^l a_i b_i - \sum_{i=l+1}^n a_i & B = \Theta \\ 0 & \text{其他} \end{cases}$$

以上讨论了两层证据的处理，对于多层证据可依次类推。

多层证据处理也可以看成是第七章 §5 折扣概念的推广。在第七章中，证据的不确定性用折扣率  $\alpha$  来表达，其中  $0 \leq \alpha \leq 1$ 。如果证据所具有的不确定性使折扣率为  $\alpha$ ，那么，除  $\Theta$  外所有焦元的基本可信度分配都折扣  $1 - \alpha$ 。在本节中，证据的不确定性用另外的证据在该证据所对应的识别框架上产生的基本可信度分配来表达。这里也可以说存在着折扣的问题。但是折扣数却是不相同的，有的折扣的多，有的折扣的少，有的甚至没有折扣，即证据的不确定性对所有焦元（推而广之，对所有命题）的影响是不一样的，这一点似乎比第七章 §5 的折扣概念更加符合人们的直观理解。另外，第七章 §5 的折扣是单层折扣，这里的折扣却是多层折扣。

前面说过，多层证据处理可陈述为反向推理的形式，因此多层证据处理有可能在专家系统中得到应用。

### §3 无限框架上证据的近似处理

Shafer 在[2]中利用 Choquet 在[18]中所提出的可交类上的 $\infty$ 阶单调函数的概念给出了一个无限框架 $\Omega$ 上的信度函数的定义,但是对于该信度函数如何处理,却没有给出什么方法。本节利用 Z. Pawlak 1982 年提出的粗糙集的概念,<sup>[9]</sup>将无限框架上的证据处理转化为有限框架上的证据处理,从而使无限框架上的证据处理有了一个比较可行的实用方法。

#### 一、无限框架上信度函数的定义

**定义 8.3.1** 对于一个集合 $\Omega$ 的子集类 $\mathcal{G}$ ,如果 $\forall A, B \in \mathcal{G}$  都有 $A \cap B \in \mathcal{G}$ ,即 $\mathcal{G}$ 对集合交封闭,则称 $\mathcal{G}$ 为 $\mathcal{P}(\Omega)$ 的可乘子类(multiplicative Subclass)。

**定义 8.3.2** 设 $\mathcal{G}$ 为 $\mathcal{P}(\Omega)$ 的可乘子类, $g$ 为定义在 $\mathcal{G}$ 上的实值函数,即 $g: \mathcal{G} \rightarrow \mathbb{R}$ 。若 $\forall A \in \mathcal{G}, A_1 \in \mathcal{G}, \dots, A_n \in \mathcal{G}$  且 $A \supset A_1, \dots, A \supset A_n$  都有

$$g(A) \geq \sum_{\emptyset \neq I \subseteq \{1, \dots, n\}} (-1)^{|I|+1} g\left(\bigcap_{i \in I} A_i\right)$$

则 $g$ 称为 $n$ 阶单调的;又若 $g$ 对所有 $n \geq 1$ 都是 $n$ 阶单调的,则称 $g$ 为 $\infty$ 阶单调。

**定义 8.3.3** 设 $\mathcal{G}$ 是 $\mathcal{P}(\Omega)$ 的可乘子类, $\phi \in \mathcal{G}, \Omega \in \mathcal{G}$ 。 $f: \mathcal{G} \rightarrow \mathbb{R}$ 为定义在 $\mathcal{G}$ 上的实值函数且满足 $f(\phi) = 0, f(\Omega) = 1$ 以及 $\infty$ 阶单调性,则称 $f$ 为 $\mathcal{P}(\Omega)$ 的可乘子类 $\mathcal{G}$ 上的信度函数。

易见,若 $f$ 为 $\infty$ 阶单调函数,则 $f$ 必为递增函数,因此一个信度函数总取值在 $[0, 1]$ 上。

**定义8.3.4** 设  $\mathcal{G}$  是包含  $\phi$  和  $\Omega$  的  $\mathcal{P}(\Omega)$  的可乘子类,  $\mathcal{F}$  是包含  $\phi$  和  $\mathcal{X}$  的  $\mathcal{P}(\mathcal{X})$  的可乘子类。若映射  $r: \mathcal{G} \rightarrow \mathcal{F}$  满足

- (1)  $r(\phi) = \phi$ ;
- (2)  $r(\Omega) = \mathcal{X}$ ;
- (3)  $\forall A, B \in \mathcal{G}, r(A \cap B) = r(A) \cap r(B)$ 。

则  $r$  称为  $\cap$ -同态 ( $\cap$ -homomorphisms)。

**定理8.3.1** 若  $f$  是  $\mathcal{F}$  上的一个信度函数,  $r$  是从  $\mathcal{G}$  到  $\mathcal{F}$  的  $\cap$ -同态, 那么  $f \circ r$  是  $\mathcal{G}$  上的信度函数。

**定理8.3.2** 设  $\mathcal{G}$  为  $\mathcal{P}(\Omega)$  的一个可乘子类,  $f$  为  $\mathcal{G}$  上的信度函数。则存在一个集合  $\mathcal{X}$ ,  $\mathcal{X}$  的子集的一个代数  $\mathcal{F}$ ,  $\mathcal{F}$  上的一个有限可加概率测度  $\mu$  和一个  $\cap$ -同态  $r: \mathcal{G} \rightarrow \mathcal{F}$ , 使  $f = \mu \circ r$ 。

**定义8.3.5** 与一个正的完全可加概率测度相联系的完备的布尔代数称为一个概率代数, 用  $\mathcal{M}$  来表示。所谓  $\mathcal{M}$  上的测度  $\mu$  是正的指的是: 对  $\mathcal{M}$  的每一个非零元素  $M$ , 都有  $\mu(M) > 0$ ;  $\mu$  为完全可加的指的是: 只要  $\mathcal{B}$  是  $\mathcal{M}$  的两两不交的元素汇集就有  $\mu(\bigvee \mathcal{B}) = \sum_{M \in \mathcal{B}} \mu(M)$  ( $\bigvee \mathcal{B}$  意思是对  $\mathcal{B}$  中所有元素求并)。

**定理8.3.3** 设  $f$  是可乘子类  $\mathcal{G}$  上的信度函数, 那么存在一个概率代数  $\mathcal{M}$  以及与  $\mathcal{M}$  相联系的测度  $\mu$  和一个  $\cap$ -同态  $\rho: \mathcal{G} \rightarrow \mathcal{M}$ , 使  $f = \mu \circ \rho$ 。交同态  $\rho$  称为一个概率分配 (Allocation probability)。

该定理说明, 对每一信度函数一定存在一个概率代数使得该信度函数在该代数里能用一个概率分配来表示。

定理 8.3.2 和定理 8.3.3 是 Shafer 在[2]中所提出的两个最深刻的定理。它们说明了一个框架上的信度函数可以转化为另一框架上的一个概率测度来研究。但是,对于如何实现这种转化以及一个无限框架上的信度函数能否转化为一个有限框架上的信度函数(当然概率测度也好!)来研究这样两个问题没有给予回答,这说明应用 Shafer 的这种理论仍然不能处理无限框架上的证据,至少也可以说这种理论没有为我们提供一种无限框架上证据处理的实用方法。本文导出的方法可以使无限框架上的信度函数近似地转化为一个有限框架上的信度函数来处理,从而使实践中经常出现的无限框架上的证据处理问题有可能得以解决。

Shafer 在[2]中还讨论了定义在  $\mathcal{P}(\Omega)$  的可乘子类  $\mathcal{E}$  上的信度函数向  $\mathcal{P}(\Omega)$  的扩充问题。他证明了  $\mathcal{E}$  上的信度函数  $f$  总存在一个到  $\mathcal{P}(\Omega)$  的标准扩充,即由下式给定的信度函数  $\bar{f}$

$$\bar{f}(A) = \sup \left\{ \sum_{i \neq j \in \{1, \dots, n\}} (-1)^{|I|+1} f \left( \bigcap_{i \in I} A_i \right) \right\}$$

其中,上确界是对所有  $n \geq 1$  和所有满足  $A \supset A_1, \dots, A \supset A_n$  的  $\mathcal{E}$  的元素  $A_1, \dots, A_n$  所取的。

这种扩张之所以称作是标准的,是因为它是最小的,即对任何由  $f$  扩张来的  $\mathcal{P}(\Omega)$  上的其他信度函数  $g$  来讲

$$\bar{f} \leq g$$

若  $\bar{f}$  是  $f$  在  $\mathcal{P}(\Omega)$  上的标准扩充,则也称  $f$  是  $\bar{f}$  在  $\mathcal{E}$  上的投影,记为  $f = \bar{f}|_{\mathcal{E}}$ 。

另外,Shafer 还讨论了无穷集  $\Omega$  上的信度函数的两个

正规条件：连续性和稠密性问题等，限于篇幅，不一一列举。

## 二、粗糙集 (Rough sets) 理论

粗糙集理论是1982年由 Z.Pawlak 所创立的<sup>[8]</sup>。在这种理论中，他研究了集合上的近似运算或粗糙运算 (approximate operation 或 Roughly operation)、集合的近似相等以及近似包含。他认为这种理论是另一种模糊集理论和另一种容差理论，而且在一些人工智能的分支上可以有很重要的应用。这些分支诸如：归纳推理 (inductive reasoning)、自动分类 (automatic classification)、模式识别 (pattern recognition)、学习算法 (learning algorithms)。另外在分类理论 (classification)、聚类分析 (cluster analysis)、测量理论 (measurement theory)、分类法 (taxonomy) 中也会结出丰硕的研究和应用成果。Z.Pawlak 强调了他的这种粗糙集理论可以看成是人工智能的数学基础，而不是一种新的集合论或新的分析。

设  $U$  是一个论域 (Universe)， $R$  是论域  $U$  上的一个等价关系，则  $A = (U, R)$  称为一个近似空间 (approximation space)。 $R$  称为一个不可分辨的关系 (indiscernibility relation)。若  $x, y \in U, (x, y) \in R$ ，则  $x, y$  称为在  $A$  中是不可分辨的 (indistinguishable)。关系  $R$  的等价类将称为  $A$  的基本集或  $A$  的原子 (elementary sets 或 atoms)，简称为基本集或原子。所有原子的集将表示为  $U/R$ ，我们假定空集是所有  $A$  的基本集。 $A$  的基本集的每一个有限并称为  $A$  的组合集 (composed set) 或简单地称为组合集。 $A$  的所有组合集的族表示为  $\text{Com}(A)$ ，显然  $\text{Com}(A)$  是一个布尔代数即所有组合集的

族在集的并，交和补下是封闭的。

$U$  的子集将用  $X, Y, Z$  也可能加上一些下标来表示， $\phi$  也用 0 表示， $U$  也用 1 表示。

设  $X$  是  $U$  的一个给定的子集，包含  $X$  的  $A$  的最小的组合集称为  $X$  在  $A$  中的上近似 (upper approximation)，用符号  $\bar{A}(X)$  表示。

显然， $\bar{A}(X) = \{x_i[x]_R \cap X \neq \phi\}$ 。

包含在  $X$  中的最大的组合集称为  $X$  在  $A$  中的下近似 (lower approximation)，用符号  $\underline{A}(X)$  来表示。

显然， $\underline{A}(X) = \{x_i[x]_R \subset X\}$ 。

对这样定义的近似运算，下列特性成立

$$\bar{A}(X) \supset X \supset \underline{A}(X)$$

$$\bar{A}(1) = \bar{A}(0) = 1 \quad (\text{该式说明 } A \text{ 的所有原子的个数有限, 即在这种理论中 } R \text{ 的等价类的个数要求有限})$$

$$\underline{A}(0) = \bar{A}(0) = 0$$

$$\bar{A}(\bar{A}(X)) = \underline{A}(\bar{A}(X)) = \bar{A}(X)$$

$$\underline{A}(\underline{A}(X)) = \bar{A}(\underline{A}(X)) = \underline{A}(X)$$

$$\bar{A}(X \cup Y) = \bar{A}(X) \cup \bar{A}(Y)$$

$$\underline{A}(X \cap Y) = \underline{A}(X) \cap \underline{A}(Y)$$

$$\bar{A}(X) = -\underline{A}(-X)$$

$$\underline{A}(X) = -\bar{A}(-X)$$

其中， $-X$  是  $U - X$  的简化。

设  $A = (U, R)$  是近似空间， $X, Y \subset U$ ，如果  $\underline{A}(X) = \underline{A}(Y)$ ，则称  $X, Y$  在  $A$  中粗糙下相等 (Roughly bottom equal)，用符号  $X \approx_{\underline{A}} Y$  表示。



如果  $\bar{A}(X) = \bar{A}(Y)$ , 则称  $X, Y$  在  $A$  中粗糙上相等 (Roughly top-equal), 用符号  $X \simeq_A Y$  表示。

如果  $X \approx_A Y$  且  $X \simeq_A Y$ , 则称  $X, Y$  在  $A$  中粗糙相等 (Roughly equal), 用符号  $X \approx_A Y$  表示。

类似地也可以导出粗糙 (上, 下) 包含的概念。

粗糙 (上, 下) 相等和粗糙 (上, 下) 包含有许多特性, 在此不一一列举, 见[9]。

另外在[9]中 Z. Pawlak 根据上述相等和包含的概念还给出粗糙集的概念。一个粗糙 (上, 下) 集是所有粗糙 (上, 下) 相等的集合的族, 该族关于关系  $\approx_A (\simeq_A, \simeq_A)$  是等价的。关于粗糙集的特性也不再详述, 请参看[9]。

### 三、无限框架上信度函数的逼近

设  $\Omega$  是一个无穷集,  $\mathcal{S}$  是  $\mathcal{P}(\Omega)$  的包含  $\phi$  和  $\Omega$  的可乘子类,  $f$  是定义在  $\mathcal{S}$  上的信度函数 (对于定义在  $\mathcal{P}(\Omega)$  上的信度函数处理方法更加简单)。现在设想在集合  $\Omega$  上存在一个不可分辨的等价关系  $R$ ,  $R$  的等价类有限, 则由此形成了一个近似空间  $A = (\Omega, R)$ 。

设  $\Theta = \Omega/R$ , 即  $\Theta$  为  $A$  的所有原子的集合, 不妨设  $\Theta = \{\theta_1, \dots, \theta_n\}$ , 则  $\mathcal{P}(\Theta)$  与  $\text{Com}(A)$  形成了一一对应 (因为, 若  $\theta'_1,$

$\dots, \theta'_n \in \Theta$ , 则  $\{\theta'_1, \dots, \theta'_n\} \leftrightarrow \bigcup_{i=1}^n \theta'_i$ )

**定理 8.3.4** 设  $A, B \subset \Theta$

$$A = \{\theta_1^1, \theta_2^1, \dots, \theta_j^1\}$$

$$B = \{\theta_1^2, \theta_2^2, \dots, \theta_i^2\}$$

$$A \cap B = \{\theta_1^1, \theta_2^1, \dots, \theta_i^1\}$$

$$\text{则} \quad A \cap B \leftrightarrow \left( \bigcup_{k=1}^j \theta_k^1 \right) \cap \left( \bigcup_{k=1}^i \theta_k^2 \right)$$

$$\text{即} \quad \bigcup_{k=1}^j \theta_k' = \left( \bigcup_{k=1}^j \theta_k^1 \right) \cap \left( \bigcup_{k=1}^i \theta_k^2 \right) \quad (8.3.1)$$

证:

$$(1) \quad \forall x \in \bigcup_{k=1}^j \theta_k \quad \text{则 } x \text{ 属于某一个 } \theta_k'$$

$$\because \theta_k' \in A \cap B \quad \therefore \theta_k' \in A \text{ 且 } \theta_k' \in B$$

$$\therefore x \in \bigcup_{k=1}^j \theta_k^1 \text{ 且 } x \in \bigcup_{k=1}^i \theta_k^2$$

$$\therefore x \in \left( \bigcup_{k=1}^j \theta_k^1 \right) \cap \left( \bigcup_{k=1}^i \theta_k^2 \right)$$

$$\therefore \bigcup_{k=1}^j \theta_k' \subset \left( \bigcup_{k=1}^j \theta_k^1 \right) \cap \left( \bigcup_{k=1}^i \theta_k^2 \right)$$

$$(2) \quad \forall x \in \left( \bigcup_{k=1}^j \theta_k^1 \right) \cap \left( \bigcup_{k=1}^i \theta_k^2 \right)$$

$$\text{则} \quad x \in \bigcup_{k=1}^j \theta_k^1 \text{ 且 } x \in \bigcup_{k=1}^i \theta_k^2$$

$$\therefore x \text{ 属于某一个 } \theta_k^1 \text{ 且 } x \text{ 属于某一个 } \theta_k^2$$

不妨假设  $x \in \theta_k^1$  且  $x \in \theta_k^2$

其中  $\theta_k^1 \in A$   $\theta_k^2 \in B$

$$\therefore \theta_k^1 \cap \theta_k^2 \neq \emptyset$$

又  $\because \theta_k^1, \theta_k^2$  是  $R$  的等价类

$\therefore \theta_k^1 = \theta_k^2$  不妨设为  $\theta_k'$

$\therefore \theta_k' \in A \cap B$

$\therefore x \in \theta_k'$

$\therefore x \in \bigcup_{k=1}^l \theta_k'$

$\therefore \bigcup_{k=1}^l \theta_k \supset (\bigcup_{k=1}^l \theta_k^1) \cap (\bigcup_{k=1}^l \theta_k^2)$

证毕

设  $\tilde{f}$  是  $f$  到  $\mathcal{P}(\Omega)$  的标准扩充, 则  $\tilde{f}|_{\mathcal{S}} = f$ 。

以下先讨论  $\tilde{f}$  的处理问题。

对  $\forall X \subset \Omega$ ,

$$\underline{A}(X) = \{x; [x]_a \subset X\} \in Com(A)$$

由此可建立映射  $r: \mathcal{P}(\Omega) \rightarrow \mathcal{P}(\Theta)$  满足

(1)  $r(\phi) = \phi$ ;

(2)  $r(\Omega) = \Theta$ ;

(3)  $\forall X \in \mathcal{P}(\Omega), r(X) = \{\theta_1', \dots, \theta_k'\}$ ,

其中,  $\theta_1', \dots, \theta_k'$  满足  $\underline{A}(X) = \bigcup_{i=1}^k \theta_i'$ 。

由下面定理可知, 这样定义的映射  $r$  为一个  $\cap$ -同态。

**定理 8.3.5** 上面定义的映射  $r$  为一个  $\cap$ -同态。

证: 只要证

$r$  满足  $\forall X, Y \in \mathcal{P}(\Omega)$

$$r(X \cap Y) = r(X) \cap r(Y)$$

即可!

设  $\underline{A}(X) = \bigcup_{k=1}^l \theta_k^1$   $\underline{A}(Y) = \bigcup_{k=1}^j \theta_k^2$

其中,  $\theta_1^1, \dots, \theta_l^1$  和  $\theta_1^2, \dots, \theta_l^2$  都是  $\Theta$  的元素, 则由  $r$  的定义知

$$r(X) = \{\theta_1^1, \dots, \theta_l^1\}$$

$$r(Y) = \{\theta_1^2, \dots, \theta_l^2\}$$

又由下近似的性质  $\underline{A}(X \cap Y) = \underline{A}(X) \cap \underline{A}(Y)$  知:

$$\underline{A}(X \cap Y) = \left( \bigcup_{k=1}^l \theta_k^1 \right) \cap \left( \bigcup_{k=1}^l \theta_k^2 \right)$$

设  $r(X) \cap r(Y) = \{\theta_1', \dots, \theta_l'\}$

则由(8.3.1)式知

$$\left( \bigcup_{k=1}^l \theta_k^1 \right) \cap \left( \bigcup_{k=1}^l \theta_k^2 \right) = \bigcup_{k=1}^l \theta_k'$$

$$\therefore \underline{A}(X \cap Y) = \bigcup_{k=1}^l \theta_k'$$

$\therefore$  由定义知:  $r(X \cap Y) = \{\theta_1', \dots, \theta_l'\}$

$\therefore r(X \cap Y) = r(X) \cap r(Y)$

证毕

下面我们根据  $\bar{f}$ , 在  $\Theta$  上构造一个信度函数。

**定理8.3.6**  $\forall A = \{\theta_1', \dots, \theta_k'\} \subset \Theta$ ,

令  $\mu(A) = \bar{f}\left(\bigcup_{i=1}^k \theta_i'\right)$ , 则  $\mu: \mathcal{P}(\Theta) \rightarrow [0, 1]$  为一信度函数。

证: (1)  $\mu(\Phi) = \bar{f}(\Phi) = 0$

$$(2) \mu(\Theta) = \bar{f}\left(\bigcup_{i=1}^n \theta_i\right) = \bar{f}(\Omega) = 1$$

(3)  $\forall A_1, \dots, A_l \subset \Theta$ , 不妨设

$$A_1 = \{\theta_1^1, \theta_2^1, \dots, \theta_{k_1}^1\}$$

$$A_2 = \{\theta_1^2, \theta_2^2, \dots, \theta_{\lambda_2}^2\}$$

.....

$$A_l = \{\theta_1^l, \theta_2^l, \dots, \theta_{\lambda_l}^l\}$$

若  $I \subset \{1, \dots, l\}$ , 再令  $\bigcap_{j \in I} A_j = \{\theta_1^I, \theta_2^I, \dots, \theta_{\lambda_I}^I\}$

$$\text{则 } \mu\left(\bigcup_{i=1}^l A_i\right) = \mu(\{\theta_1^1, \dots, \theta_{\lambda_1}^1, \theta_1^2, \dots, \theta_{\lambda_2}^2, \dots, \theta_1^l, \dots, \theta_{\lambda_l}^l\})$$

$$= j\left(\bigcup_{j=1}^l \left(\bigcup_{i=1}^{\lambda_j} \theta_i^j\right)\right)$$

$$\geq \sum_{\emptyset \neq I \subset \{1, \dots, l\}} (-1)^{|I|+1} j\left(\bigcap_{j \in I} \left(\bigcup_{i=1}^{\lambda_j} \theta_i^j\right)\right)$$

( $\because j$  为信度函数)

$$= \sum_{\emptyset \neq I \subset \{1, \dots, l\}} (-1)^{|I|+1} j\left(\bigcup_{k=1}^{\lambda_I} \theta_k^I\right)$$

(由 (8.3.1) 式)

$$= \sum_{\emptyset \neq I \subset \{1, \dots, l\}} (-1)^{|I|+1} \mu(\{\theta_1^I, \theta_2^I, \dots, \theta_{\lambda_I}^I\})$$

(由  $\mu$  的定义)

$$= \sum_{\emptyset \neq I \subset \{1, \dots, l\}} (-1)^{|I|+1} \mu\left(\bigcap_{j \in I} A_j\right)$$

(由假设)

$\therefore \mu: \mathcal{P}(\Theta) \rightarrow [0, 1]$  为一信度函数。证毕

构造信度函数  $\mu$  以后, 根据定理 8.3.1 知:  $\mu \circ \tau$  也为  $\mathcal{P}(\Omega)$  上的信度函数, 记为  $I$ 。

$\forall X \in \mathcal{P}(\Omega)$ , 设  $\underline{A}(X) = \bigcup_{i=1}^j \theta_i$  中,  $\theta_1, \dots, \theta_j$  是  $\Theta$  中

的元素, 则

$$\begin{aligned} r(X) &= \{\theta_1, \dots, \theta_j\} \\ \therefore \underline{f}(X) &= \mu \circ r(X) = \mu(r(X)) \\ &= \mu(\{\theta_1, \dots, \theta_j\}) \\ &= \underline{f}\left(\bigcup_{i=1}^j \theta_i\right) \\ &= \underline{f}(\underline{A}(X)) \end{aligned}$$

所以说, 如果  $X \approx_A Y$ , 则  $\underline{A}(X) = \underline{A}(Y)$ ,  
从而  $\underline{f}(X) = \underline{f}(Y)$ 。

因此, 若用  $\underline{f}$  近似  $\tilde{f}$  的话, 则所有粗糙下相等的集合的信度都等于该集合的下近似的信度。直观上对于粗糙下相等的两个集合, 它们的信度应该相差不大, 所以在  $\Omega$  上用  $\underline{f}$  近似  $\tilde{f}$  是合理的。因为  $\tilde{f}/\mathcal{E} = f$ , 所以在  $\mathcal{E}$  上用  $\underline{f}/\mathcal{E}$  近似  $f$  也就是合理的。

这样, 无穷框架  $\Omega$  上的证据处理的整个过程可以概述为

(1) 在该框架上寻找一个不可分辨的等价关系  $R$ , 由此形成近似空间  $A = (\Omega, R)$ , 从而构造出有限框架  $\Theta$ ;

(2) 根据框架  $\Omega$  上的信度函数  $f$ , 按上述方法诱导一个  $\Theta$  上的信度函数  $\mu$ ;

(3) 处理  $\Theta$  上的信度函数  $\mu$  (可按 Shafer 提供的有限框架上的信度函数的所有处理方法进行)。

(4) 通过  $\Pi$ -同态  $r$  (由  $\Omega$  的集合的下近似定义) 把  $\Theta$  上的处理结果重新送回框架  $\Omega$ , 近似代替  $\Omega$  上的处理结

果。

比如说，设  $f_1$  和  $f_2$  是定义在  $\mathcal{P}(\Omega)$  上的两个信度函数，那么  $f_1$  与  $f_2$  如何合成呢？

利用本节所述方法，可以求得  $f_1$  与  $f_2$  的合成中的一个近似。

设在  $\Omega$  上有一个等价关系  $R$ ，则由此可构造有限框架  $\Theta$  以及  $\cap$ —同态  $r$ 。如果  $\mu_1$  是由  $f_1$  按上法所诱导的  $\Theta$  上的信度函数， $\mu_2$  是  $f_2$  按上法所诱导的  $\Theta$  上的信度函数，则在  $\Theta$  上可求得  $\mu_1 \oplus \mu_2$ 。然后用  $r$  返回，即求  $(\mu_1 \oplus \mu_2) \circ r$ 。用它即可近似代替  $f_1$  与  $f_2$  的合成  $f_1 \oplus f_2$ 。

用这种方法近似处理无穷框架  $\Omega$  上的信度函数，一个首要的问题就是等价关系  $R$  的选择。

在进行证据处理时，人们对两个相差不大的集合（两个差不多的命题）的信度应该相差不大，换句话说，当一个命题稍微发生一点变化时，我们的信度基本上不变或改变很少，不可能出现急聚变化的情况。所以  $R$  的选择可以依赖于实际问题，将那些信度相差不大而且命题又非常相似的集的元素放到一块作为一个等价类。

另外，在框架  $\Omega$  上寻找等价关系构造框架  $\Theta$ ，当  $\Omega$  有限时等价于框架的粗化，但当  $\Omega$  无限时与框架的粗化是两个完全不同的概念。不过这种方法却提醒我们去考虑无穷框架的转化问题（粗化和细分以及相容）。只有解决了无穷框架的转化问题，无穷框架上的证据处理才能在理论上得到完整。

## 第九章 利用Shafer证据理论 作决策的一般方法

### §1 问题的提出

设有一个判决问题，对于该问题我们所能认识到的可能结果用集合 $\Theta$ 表示。设有一批证据 $E$ ，这批证据使我们在 $\Theta$ 上产生了一个基本可信度分配，即我们以多大的可能性认为 $\Theta$ 的某个子集包含该判决问题的真值。基本可信度分配的得到是证据 $E$ 在 $\Theta$ 上作用的结果，换句话说基本可信度分配是证据 $E$ 在 $\Theta$ 上的体现，在 $\Theta$ 上证据 $E$ 可以用基本可信度分配来表达。所以，对一个判决，如果有一批证据的话，那么在框架 $\Theta$ 上得到一个基本可信度分配。如果有好几批证据可利用，则我们就可以得到好几个基本可信度分配。这几个基本可信度分配是我们作出该判决的基础。应用Shafer证据理论，得到几个基本可信度分配以后，可以求出它们的合成，用来表示其对应证据的联合作用；求出合成以后，可以进一步求出一个信度函数（或似真函数，众信度函数）。在Shafer证据理论中，该信度函数就是我们判决的结果。所以利用Shafer证据理论作判决，所得到的结果是一些集合，而不是单点。即判



决的结果是我们认为真值在某个集中有多大可能，在另外的集中又有多大可能。可是实际问题却要求我们对真值究竟是多少作出一个确切的回答。从这个意义上讲，对于判决或更一般地讲对于决策，Shafer证据理论才作了一半的工作。这一章，我们要根据由证据理论所得到的信度函数进一步推出真值究竟是多少（至少也要将真值限制在最小的范围内），使信度理论能真正用于决策的制定。

## §2 用“最少点”原则求真值

上面说过，得到信度函数  $Bel$  以后，事实上我们得到了一个一个的集合，以及我们对该集合包含给定的真值的信度。正是由于我们得到的是集合而不是单点，所以我们才不能确切地知道真值。因此我们希望能通过得到的集合及其信度值找到真值，或即使找不到确切的真值，能否将所得到的集合再缩小缩小，使得包含真值的集合元素个数达到最少。这一节我们的目的就是探讨如何利用这样的“最少点”原则缩小给定的集合，以最终向真值逼近。

设我们得到的集合是  $A$ ，信度是  $Bel(A)$ 。我们希望缩小该集合，即从该集合中去掉若干个元素。但是，这个过程不能一下子进行，而应该逐步进行。否则的话，由于去掉的元素太多，信度可能会发生很大的改变，甚至会出现这种情况，即如果去掉的元素中恰好有真值，那么剩下元素的集合的信度就可能降得很低，有时甚至可能降到0。这样的话就起不到找出真值的作用。那么去掉多少元素合适呢？我们的原则是：设对于集合  $A$ ，去掉若干个元素以后的集合为  $B$ ，信度为

$\text{Bel}(B)$ 。如果 $\text{Bel}(A)$ 与 $\text{Bel}(B)$ 相差不大,说明去掉这几个元素后,对我们的信度影响不大,那么可以去掉这几个元素。

为了防止信度发生突变,我们不妨先考虑去掉一个元素的情况。

设集合 $A$ 的信度为 $\text{Bel}(A)$ ,  $A$ 分别去掉1个单点的各个子集记为 $A_1, \dots, A_K$ , 求出使 $\text{Bel}(A_i)$  ( $i=1, \dots, K$ )最大的 $i$ , 记为 $J$ 。然后比较 $\text{Bel}(A)$ 与 $\text{Bel}(A_J)$ 。如果两者相差不大, 则可以去掉那个元素, 用 $A_J$ 代替 $A$ ; 如果两者相差很大, 则说明不能去掉那个元素, 不能用 $A_J$ 代替 $A$ 。此处 $A_J$ 在 $A_1, \dots, A_K$ 中信度最大, 说明在 $A$ 中去掉 $A_J$ 对应的那个元素更有可能, 如果该元素都去不掉, 则其它元素更不能考虑。

用 $A_J$ 代替 $A$ 以后, 对 $A_J$ 同样可以考虑这样的问题, 即先求分别去掉 $A_J$ 的一个单点的各子集, 然后找出诸子集中信度最大者与 $A_J$ 的信度作比较。这个过程一直进行下去直到某个子集 $B$ 不能用分别去掉一个单点所得到的各子集中信度最大者 $B_J$ 来代替为止。 $B$ 不能用 $B_J$ 来代替说明在这批证据下, 我们所能作出的最好决策就是子集 $B$ 了。如果还想缩小 $B$ 的话, 就必须重新寻找新的证据。

在以上这些讨论中, 我们没有考虑在 $A$ 分别去掉一个单点所得到的诸子集 $A_1, \dots, A_K$ 中信度最大者有许多个的情况。如果 $A_{J_1}, \dots, A_{J_l}$ 的信度都在 $A_1, \dots, A_K$ 中达到最大, 那么我们又该如何处理呢?

$A_{J_1}, \dots, A_{J_l}$ 的信度在 $A_1, \dots, A_K$ 中达到最大, 说明除了 $A_{J_1}, \dots, A_{J_l}$ 的公共部分以外( $l \neq K$ 时 $A_{J_1}, \dots, A_{J_l}$ 的交集不会为空) 其余部分的改变不影响信度变化。所以我们可以考虑去掉 $A_{J_1}, \dots, A_{J_l}$ 的公共部分以外的部分。此时我们有

可能去掉多个元素。

例如, 设  $A = \{1, 2, 3, 4, 5\}$ , 则  $A$  分别去掉一个单点所得到的诸子集为

$$A_1 = \{1, 2, 3, 4\}$$

$$A_2 = \{1, 2, 3, 5\}$$

$$A_3 = \{1, 2, 4, 5\}$$

$$A_4 = \{1, 3, 4, 5\}$$

$$A_5 = \{2, 3, 4, 5\}$$

设  $A_1, A_3, A_4$  的信度在  $A_1, A_2, A_3, A_4, A_5$  中达到了最大。由于  $A_2 \cap A_3 \cap A_4 = \{1, 5\}$ , 所以  $A_2, A_3, A_4$  的信度相同(最大), 说明在集合  $\{1, 5\}$  上添加  $\{2, 3\}$  或者  $\{2, 4\}$  或者  $\{3, 4\}$  都不影响信度的改变。因此在  $A_2, A_3, A_4$  中我们可以考虑去掉  $\{2, 3\}, \{2, 4\}$  以及  $\{3, 4\}$ 。

那么究竟能否去掉公共部分以外的部分呢? 对一般情况,

我们假设  $C = \bigcap_{i=1}^l A_{j_i}, D = \bigcup_{i=1}^l (A_{j_i} - C)$ , 则我们考虑,

去掉的部分就是  $D$ 。为了决定能否去掉  $D$ , 我们考查  $\text{Bel}(A)$  与  $\text{Bel}(C)$  的差异。如果它们相差不大, 说明在  $A$  中去掉  $D$  以后对我们的信度影响不大, 可以去掉  $D$ 。否则说明  $\text{Bel}(A)$  与  $\text{Bel}(C)$  相差太大, 即去掉  $D$  以后对我们的信度影响太大, 这样的话, 我们就有理由认为在去掉的  $D$  中可能包含有真值, 因此我们反而可以考虑保留  $D$  而去掉  $C$ 。那么究竟能否去掉  $C$  呢? 与上相同我们仍需考查  $\text{Bel}(A)$  与  $\text{Bel}(D)$  的差异。如果相差不大, 说明可以去掉; 如果相差太大, 说明不能去掉。

值得说明的是: 以上过程必须从  $\Theta$  开始, 逐步去掉可以去掉的元素。另外, 该过程也不是一步能完成的, 而必须经

过许多步递归以后才能完成。这一点在下节的一般算法中  
可以看得更加清楚。

### §3 利用Shafer证据理论 作决策的一般算法

本节利用上节的“最少点”原则得到了一个利用 Shafer  
证据理论作决策的一般算法。

Step1 输入识别框架 $\Theta$ 及所获得的基本可信度分配

$$m_1, m_2, \dots, m_n$$

step2 IF  $n=1$  THEN BEGIN  $m = m_1$   
GOTO step4  
END

step3 利用 Dempster 合成法则求  $m_1, \dots, m_n$  的直和

$$\forall A \subset \Theta \quad m(A) = \frac{\sum_{B \cap C \cap \dots \cap F = A} m_1(B) m_2(C) \dots m_n(F)}{\sum_{B \cap C \cap \dots \cap F \neq \emptyset} m_1(B) m_2(C) \dots m_n(F)}$$

step4 根据上面得到的  $m$  求 Bel

$$\forall A \subset \Theta \quad \text{Bel}(A) = \sum_{B \subset A} m(B)$$

step5  $A = \Theta$ , CALL MAXBEL( $A, B$ )然后打印  $B$

step6 IF 又来了一批证据, 即又得到一个基本可信  
度分配  $m'$

THEN BEGIN  $m_1 = m \quad m_2 = m'$   
GOTO STEP3

END

```

step7 stop
procedure MAXBEL(A,B)
step1 求  $A$  分别去掉一个单点的各子集, 记为  $A_1, \dots, A_k$ 
step2 求使  $\text{Bel}(A_i)$  最大的  $i$ , 记为  $J_1, \dots, J_l$ 
step3 IF  $l = k$  THEN BEGIN
         $B = A$ 
        RETURN
    END

step4 求  $C = \bigcap_{i=1}^l A_{J_i}$ 
Step5 IF  $\text{Bel}(A) - \text{Bel}(C) < \varepsilon$  THEN
    BEGIN
        IF  $\text{Bel}(C) < p$  THEN BEGIN
             $B = C$ 
            RETURN
        END
    ELSE BEGIN
        IF  $|C| = 1$  THEN BEGIN
             $B = C$ 
            RETURN
        END
    ELSE BEGIN
         $A = C$ 
        GOTO STEP 1
    END
    END
END

```

```

step6 求  $D = \bigcup_{i=1}^I (A_{J_i} - C)$ 

Step7 IF  $\text{Bel}(A) - \text{Bel}(D) < \varepsilon$  THEN
    BEGIN
        IF  $\text{Bel}(D) < P$  THEN BEGIN
            B = D
            RETURN
        END
    ELSE BEGIN
        IF  $|D| = 1$  THEN BEGIN
            B = D
            RETURN
        END
    ELSE BEGIN
        A = D
        GOTO STEP1
    END
END

END

step8 B = A
step9 RETURN

```

用过程 MAXBEL 可以求到包含的元素个数最少，而信度相对来讲又最大的集合。由上节讨论可知，该集合就是相对这批证据，我们所能取的合适的决策集。在该过程中，一个集合在去掉一个元素所得到的诸子集中，信度最大的那个子集与原集合信度的接近程度用某个水平  $\varepsilon$  来刻画。如果它们的差小于  $\varepsilon$ ，说明很接近，否则的话，说明相差很大。其中  $\varepsilon$  的选取要根据所研究的问题预先给定。在该过程中，我

们给出了一个信度的下限  $P$ 。如果在缩小集合的过程中发现信度降到  $P$  以下，那么我们就应该立即停止这种缩小过程，否则我们的信度到后来可能降到很低而不能接受。下限  $P$  的选取也要根据所研究的问题预先给定。另外，在算法实现的过程中  $\epsilon$  与  $P$  的值可以随时调整，其目的就是最后找到满意的可信的决策集。

该算法在实现时会遇到一定的困难。比如，在主程序 Step4 中要求出  $Bel(A)$ ，就要首先求出  $A$  的所有子集。如果  $A$  有  $n$  个元素，那么  $A$  的所有子集的个数就为  $2^n$  个。不用说  $n$  很大，就是当  $n=10$  时， $2^n = 1024$  数值也是相当可观的，一般的计算机很难容纳，而且随着  $n$  的增大， $2^n$  成指数地增长。所以在 Step4 的计算中我们遇到的困难是 NP 问题。所以，这个算法在目前的计算机上不容易实现。但是，如果我们遇到的实际问题比较特殊的话，那么我们就可以简化  $Bel(A)$  的计算（事实上，这种情况在实际问题中是普遍存在的）。关于  $Bel$  的计算以及 Dempster 合成法则的计算问题可参看 [7]，[8] 等。

过程 MAXBEL 是一个近似求真值的过程，该过程始终是从集函数的角度来考虑的，即每次判断集  $A$  及分别去掉一个单点以后所得集中信度最大者，其信度是否相差不大。所得到的  $B$  能够满足点相对来讲最少，而信度相对来讲又最大的原则。在这批证据下，我们得到的  $B$  是最好的决策集。

但是，有些实际问题（比如说第十二章）却要求我们确切回答出真值究竟是什么。在这种情况下，我们可以针对所得到的  $B$  在  $B$  中求单点似真度最大的点。如果能求到这样的点的话，那么我们就取该点为真值（当然以一定的可信度）。

否则（比如说单点似真度最大的点有许多个）说明由  $B$  不能找到确切真值。怎么办呢？只能重新寻找证据，直到找到真值为止。

当然，要找出真值，我们也可以针对原来的  $\Theta$ ，不经过过程 MAXBEL 求  $B$ ，直接求  $\Theta$  中单点似真度最大的点作为我们的真值。可是，遗憾的很， $B$  中单点似真度最大者与  $\Theta$  中单点似真度最大者可能是不一致的。例如，设  $\Theta = \{x_1, x_2, \dots, x_{100}\}$ ， $\Theta$  上的基本可信度分配  $m$  如下所示（左边一列为焦元，右边一列为其对应的基本可信数）

$\{x_1\}$	$\frac{1}{100}$
$\{x_1, x_2\}$	$\frac{1}{100}$
$\vdots$	
$\{x_1, x_2, \dots, x_{40}\}$	$\frac{1}{100}$
$\{x_{41}\}$	$\frac{20}{100}$
$\{x_{42}, \dots, x_{80}\}$	$\frac{19}{100}$
$\{x_{81}, \dots, x_{100}\}$	$\frac{19}{100}$
$\{x_1, x_2, \dots, x_{100}\}$	$\frac{2}{100}$

对该基本可信度分配  $m$  所对应的信度函数 Bel，设  $\varepsilon = 0.04$ ， $P = 50/100$ ，用过程 MAXBEL 可得集合  $B =$



$\{x_{i1}, x_{i2}, \dots, x_{i00}\}$ 。B中单点似真度最大者为 $x_{i1}$  ( $pl(\{x_{i1}\}) = 22/100$ )。如果针对 $\Theta$ 求单点似真度最大的点则为 $x_1$  ( $pl(\{x_1\}) = 42/100$ )。

为什么会出现这种情况呢？其原因就是，直接求 $\Theta$ 中单点似真度最大的点作为我们的真值，事实上是将 $pl(\{x_i\})$  ( $i = 1, \dots, q$ )，设 $\Theta = \{x_1, \dots, x_q\}$ 按从小到大的顺序排队，用点函数 $pl(\{x_i\})$  ( $i = 1, \dots, q$ )近似代替集函数Bel或对应的pl了。一般情况下，点函数 $pl(\{x_i\})$  ( $i = 1, \dots, q$ )不能表示整个信度函数Bel或pl所包含的信息，所以由点函数 $pl(\{x_i\})$  ( $i = 1, \dots, q$ )求到的真值，一般情况下是不确切的。换句话说，似真度大的点成为真值的可能性并不一定大。

而用MAXBEL得到B然后再求B中单点似真度最大者作为真值与直接由单点似真度求真值是两种不同的近似方法，其结果当然有可能不同。

但是，对于一致支持函数用两种方法求真值，其结果却是相同的。这说明对于一致支持函数，两种近似方法是一致的。

那么，对于不是一致支持函数的信度函数，用哪种方法近似更好呢？第一种方法始终是从集函数的角度考虑的（只是到最后才从点函数的角度考虑，求B中单点似真度最大者），而第二种方法完全是从点函数的角度考虑的，因此，作者认为，第一种方法比第二种方法更精确一些。所以本书凡是遇到必须求出确切真值的情况时，总是首先用过程MAXBEL得到一个子集B，再在B中求单点似真度最大者，而不是直接求 $\Theta$ 中单点似真度最大者（参见第十二章）。

## §4 一些问题的讨论

利用Shafer证据理论作决策所依赖的是证据。那么证据究竟是什么呢？

Shafer在[1]中强调指出：在我们的理论中，证据不是某种实证据，而是我们的经验和实际知识的一部分，是我们实际从事的观察和研究的结果。

所以说，对某个实际的决策问题，我们可以凭经验和知识在框架 $\Theta$ 上得到一个基本可信度分配。但是，对于一个特定的决策者来说，他所拥有的经验和知识是有限的，是确定的，即他自己只能在框架 $\Theta$ 上产生一个基本可信度分配，依靠这个分配作决策显然是不确切的、片面的。特别是当该决策者对该决策领域的经验和知识比较贫乏时，更是如此（这种情况是经常出现的。比如说，掌握某个领域决策权的领导人是该领域的外行，而决策又必须是他拍板）。为了克服这个问题，决策者就有必要向本领域的若干专家提出咨询，将这些专家的意见综合起来，作出合适的决策。在这里，专家意见如何表达是一个值得讨论的问题。但是，最起码某一个专家可以根据它的经验和知识在框架 $\Theta$ 上产生一个基本可信度分配 $m$ ，而该专家的意见就可以用该基本可信度分配来表达（下面我们还要讨论，专家的经验用基本可信度分配来描述是最合适的，参见第十章§2）。这样的话，假设决策者向 $n$ 个专家提出了咨询，那么他就可以得到 $n$ 个基本可信度分配 $m_1, m_2, \dots, m_n$ 。这里，我们假定，假设决策者本人与某个专家具有相同的经验和知识的话，他也可以得到相同的基本可信

度分配。所以，到现在为止，我们可以认为决策者根据  $n$  批不同的证据在框架 $\Theta$ 上产生了  $n$  个不同的基本可信度分配，从而就可以利用Dempster合成法则合成这  $n$  个基本可信度分配。最后，利用合成后的基本可信度分配所对应的信度函数，根据上节算法作出决策。所以，这种决策方法可以作为一种依赖于专家咨询的方法。

在进行决策时，完全可能出现这种情况，决策集缩到最后还是很大，我们还不能找到一个合适的决策集。上面说过遇到这种情况时，说明我们的证据不够，那么为了得到新的证据，我们必须向另外的专家进行咨询，如果还不能找到合适的决策集，必须再向另外的专家进行咨询，直到找到合适的决策集为止。因此，这种方法是一种不断咨询、逐步完善的方法。

综上所述，这种决策方法的特点是：可以综合多个专家的意见，得到各个专家都普遍满意的决策结果。这为集体制定决策提供了一种确实可行的方法。

另外，对于决策者本人来讲，根据他的经验和知识在框架 $\Theta$ 上产生了一个基本可信度分配，据此可以作出一种决策。以后在实践中，随着他的经验和知识的丰富，又可以在框架 $\Theta$ 上产生新的基本可信度分配，通过与以前的基本可信度分配的合成，又可以作出一种新的决策。即通过学习，他可以随时产生新的基本可信度分配，随时作出新的决策。这正是人类学习过程的描述，所以这种决策方法为人类经验的积累和利用提供了一种可能，为用计算机模拟人的学习过程提供了一条途径。

## 第十章 决策分析

决策分析是一种适用于不确定型决策问题和风险型决策问题的决策方法，是决策理论的一个分支<sup>[21]</sup>。决策分析以概率论与数理统计为理论基础，把统计决策方法应用于不可控或不确定因素的决策问题中。决策分析以主观概率和效用函数为主要概念，是一种量化的决策方法。现在这种方法在许多领域中都得到了应用，而且取得了比较好的效果。

在决策分析中，系统状态的生存概率一般情况下不能精确地给出。因此根据人的经验、知识和对系统的了解，对系统状态生存的概率作出主观上的估计和推测（这样得到的概率称为主观概率）就是非常重要的。状态生存的主观概率给的好，即越符合客观实际，则所作出的决策就越科学，状态生存的主观概率给的不好，即越不符合客观实际，则所作出的决策理论上最优，实际上越不最优，有时甚至很差。所以研究如何科学地给出主观概率，对于决策分析方法科学地合理地用于实际的决策问题，无疑会产生巨大的推动作用。

本章§1简要地介绍了几种传统的主观概率的估计方法；§2用Shafer信度理论导出了主观概率的另一种估计方法——专家咨询法；§3导出了Dempster合成法则的递归形式，给出

了用专家咨询法得到的主观概率的修正算法；§4应用证据理论导出了另一种决策分析的方法——M决策法。

## §1 主观概率的传统估计

主观概率的传统估计方法有：等可能法，数理统计法，主观试探法，众人评估法。

等可能法认为在没有历史知识可供借鉴，难以采用任何方法对系统的生存状态进行预测的情况下，各个状态的发生、存在是等可能的。即如果状态集

$$S = \{x_1, x_2, \dots, x_q\}$$

则认为  $p(x_i) = \frac{1}{q} \quad (i = 1, 2, \dots, q)$ 。

等可能法在没有其他方法可利用时也可以采用，但是采用等可能法很难想象其决策会带有多少科学的成分。

数理统计法要求我们拥有大量的历史资料，从对历史资料的考察、分析和研究中预测未来系统状态的出现情况。如，某地自从有记载以来发生地震  $N$  次，其中轻微地震有  $n_1$  次，中等程度的地震有  $n_2$  次，强烈地震有  $n_3$  次， $n_1 + n_2 + n_3 = N$ ，据此我们当然可以认为各种地震发生的概率为  $f_i = \frac{n_i}{N} (i = 1, 2, 3)$ 。

但是，并不是所有系统都有这样的记载。事实上，能够利用这种方法求得系统的生存概率的情况是很少见的。大多数系统没有这样的记载，因而其状态的生存概率不能这样来求。还有许多系统，它的状态的出现是不可重复的，在这种

情况下,就更不能用这种方法来求状态的生存概率了,状态的生存概率需要用其他的方法进行估计。

主观试探法对状态变量  $x$  的生存概率  $p(x)$  采用比较试探的办法进行估计。首先估计  $p(x) \in \left[\frac{1}{2}, 1\right]$  还是  $\left[0, \frac{1}{2}\right]$ 。为此需要询问决策者  $x$  发生与  $x$  不发生谁更可能。如果决策者认为  $x$  发生更有可能,就认为  $p(x) \geq \frac{1}{2}$ , 如果说  $x$  不发生更有可能就认为  $p(x) \leq \frac{1}{2}$ 。现在假设确定为  $p(x) \geq \frac{1}{2}$ , 那么下面要估计

的是  $p(x) \in \left[\frac{1}{2}, \frac{3}{4}\right]$  还是  $\left[\frac{3}{4}, 1\right]$  (即将  $\left[\frac{1}{2}, 1\right]$  分成两段), 仍然采用询问的方式。这次估计作好以后, 再将得到的区间分成两段。依次下去, 直到确定出  $p(x)$  为止。

这种估计法所遇到的困难就是试探到某一步 (有时甚至是一开始) 决策者就很难断定  $p(x)$  是处在左半区间还是右半区间。为了使决策进行下去, 他不得不硬着头皮说出一种选择。所以说他的这种选择带有一种被迫性, 带有一种盲目性。这种情况当决策者对系统的认识 and 知识比较少时更加容易出现。

众人评估法是为了克服个人估计的片面性, 发挥众人智慧而设计的一种估计方法。该法首先请若干有丰富经验的专家能人, 分别对状态发生的概率作出估计, 然后由决策者将各种估计加以综合得到最后的估计结果。综合的方法是给每位专家一个权 (代表决策者对每个专家意见的信赖程度), 然后求加权平均值。

该估计法所遇到的困难是, 让专家说出某个状态发生的

概率有时是不大可能的。比如说,如果系统的状态是产品的各个不同的合格率,那么让一个老师傅说出产品的合格率是0.6的可能性有多大是很困难的。相反,如果让一个老师傅说一下产品合格率,他随口就可以说出80%在0.5以上,70%处在0.8—0.9之间等。这就是说谈论系统的状态,位于某个子集中要比谈论系统的状态究竟是什么容易的多。§2中的专家咨询法就是依据这样一个思想而设计的。

## §2 专家咨询法

上面说到决策者给出主观概率是基于他的经验、知识和对系统的了解。这些都可以看成是针对系统的各种生存状态出现的证据,统记为 $E$ 。应用Shafer证据理论,设框架 $\Theta$ 就为状态集 $S = \{x_1, \dots, x_q\}$ 。这样的话,就可以假设,决策者根据证据 $E$ 在 $\Theta$ 上产生了一个基本可信度分配 $m$ ,焦元为 $A_1, \dots, A_p$ ,可信数分别为 $m(A_1), \dots, m(A_p)$ 。

得到基本可信度分配以后,利用公式

$$pl(\{x\}) = \sum_{x \in A} m(A)$$

求出单点似真度 $pl(\{x_1\}), \dots, pl(\{x_q\})$ 。

$$\text{令 } p(x_i) = \frac{pl(\{x_i\})}{pl(\{x_1\}) + \dots + pl(\{x_q\})}$$

$$= \frac{pl(\{x_i\})}{\sum_{j=1}^q pl(\{x_j\})}$$

$$i=1, \dots, q$$

则 $p(x_1), \dots, p(x_i)$ 即可作为各种状态出现的主观概率。

事实上,上面的基本可信度分配的得到可以采用专家咨询的方法。设向 $n$ 个专家提出咨询,则每个专家都可以根据他自己的经验、知识和对系统的了解,在框架 $\Theta$ 上产生一个基本可信度分配(产生法可参照下章§2),所以在咨询结束以后,我们就可以得到 $n$ 个基本可信度分配 $m_1, \dots, m_n$ 。与上章讨论相同,这 $n$ 个基本可信度分配可以看成是决策者根据不同的证据得到的,所以可以用Dempster合成法则进行合成,合成的结果就可以作为上面所说的基本可信度分配 $m$ 。

与上章讨论相同,在利用这种专家咨询法得到主观概率估计时,体现了一种集体的力量。基于这样得到的主观概率的决策同样可以说是集体决策的结果。

人们在对状态集 $S$ 中的状态谁出现,谁不出现的估计中,往往首先说出的或容易说出的是以多大可能性出现某几种状态,而让他说出某一种状态出现不出现是困难的,有时甚至是不可能的。如:把产品废品率分为10个等级1%—10%,让一个工人师傅估计某种产品的废品率,他脱口而出的是5%以下或高于8%或2%—4%之间,而不是1%或9%。这说明我们这种主观概率的估计方法具有某种普遍性,而且是传统估计方法的推广( $A_1, \dots, A_n$ 退化为单点即为传统的估计方法)。

当然,我们的方法也不排除某个专家具体给出主观概率的情况。在这种情况下,我们可以认为这个专家所对应的证据在框架 $\Theta$ 上产生了一个基本可信度分配 $m(\{x_i\}) = p_i (i = 1, \dots, q), m(\text{其他}) = 0$ 。另外,我们的方法也不排除有统计资料可利用的情况。在这种情况下,根据统计资料得到的估



计可以看成是由证据（统计资料）所产生的一个基本可信度分配。

我们的方法也可以对所得出的主观概率进行修正，但是，这种修正决不采用贝叶斯条件法则（前几种传统的估计方法都采用贝叶斯条件法则对主观概率进行修正），而采用 Dempster 合成法则。Dempster 合成法则不管证据出现的先后，因此它比贝叶斯条件法则使用起来灵活。特别是有了 Dempster 合成法则的递归形式以后，这种修正更加方便。

### §3 Dempster 合成法则的递归形式 及主观概率的修正

设有  $n$  个信度函数，其众信度函数分别为  $Q_1, Q_2, \dots, Q_n$ ，合成以后的众信度函数为  $Q^*$ ，则由  $n$  个信度函数合成的众信度函数形式知

$$\forall A \subset \Theta \quad Q^*(A) = \frac{1}{K_n} Q_1(A) \cdot Q_2(A) \cdots Q_n(A)$$

其中 
$$K_n = \sum_{\substack{A_i \cap B_j \cap \cdots \cap C_k \neq \emptyset \\ A_i \cap B_j \cap \cdots \cap C_k \neq \emptyset}} m_1(A_i) m_2(B_j) \cdots m_n(C_k)$$

$A_i, B_j, \dots, C_k$  分别为  $Q_1, Q_2, \dots, Q_n$  的焦元

由此可知，只要知道  $K_n$  的递推关系，则也就得到了  $Q^*$  的递推关系。

下面求  $K_n$  的递推关系

$$\begin{aligned} K_n &= \sum_{\substack{A_i \cap B_j \cap \cdots \cap C_k \neq \emptyset \\ A_i \cap B_j \cap \cdots \cap C_k \neq \emptyset}} m_1(A_i) m_2(B_j) \cdots m_n(C_k) \\ &= \sum_{\substack{A_i \cap B_j \cap \cdots \cap C_k \neq \emptyset \\ A_i \cap B_j \cap \cdots \cap C_k \neq \emptyset}} m_1(A_i) m_2(B_j) \cdots m_{n-1}(E_i) m_n(C_k) \end{aligned}$$

$$= \sum_{\substack{i, j, \dots, l \\ A_i \cap B_j \cap \dots \cap E_l = A \neq \emptyset}} m_i(A_i) m_j(B_j) \dots m_{n-1}(E_l) \cdot \sum_k m_n(C_k) \\ - \sum_{\substack{i, j, \dots, l \\ A_i \cap B_j \cap \dots \cap E_l = A \neq \emptyset \\ A \cap C_k = \emptyset}} m_i(A_i) m_j(B_j) \dots m_{n-1}(E_l) \cdot m_n(C_k)$$

$$\therefore \sum_k m_n(C_k) = 1$$

$$\therefore K_n = K_{n-1} - \sum_{\substack{i, j, \dots, l \\ A_i \cap B_j \cap \dots \cap E_l = A \neq \emptyset \\ A \cap C_k = \emptyset}} m_i(A_i) m_j(B_j) \dots m_{n-1}(E_l) m_n(C_k)$$

由此我们得到了 Dempster 合成法则的递归形式，根据 Dempster 合成法则的递归形式，我们可以得到  $pl(\{x\})$  的一种递归修正算法。

$$\therefore \text{在单点上 } Q^n(\{x\}) = pl^n(\{x\})$$

$$\therefore pl^n(\{x_i\}) = \frac{1}{K_n} pl_1(\{x_i\}) pl_2(\{x_i\}) \dots pl_n(\{x_i\})$$

$$(i = 1, 2, \dots, q)$$

由此可得到一种算法如下

(1) 输入  $m_n(C_1), \dots, m_n(C_t)$ ;

(2) 计算

$$\sum_{\substack{i, j, \dots, k \\ A_i \cap B_j \cap \dots \cap E_l = A \neq \emptyset \\ A \cap C_k = \emptyset}} m_i(A_i) m_j(B_j) \dots m_{n-1}(E_l) m_n(C_k)$$

$$(\triangle K)$$

(要充分利用计算  $K_{n-1}$  时留下的信息)

(3) 计算  $K_n = K_{n-1} - K$ ;

(4) 对  $i = 1, \dots, q$

$$\text{令 } R^{n-1}(\{x_i\}) = K_{n-1} pl^{n-1}(\{x_i\})$$

$$\text{则} \quad p_i^n(\{x_i\}) = \frac{1}{K_n} K_n^{n-1}(\{x_i\}) p_i^n(\{x_i\})$$

(5) 对  $i = 1, \dots, q$  计算  $p^n(x_i)$ ;

(6) 假设又得到一个  $m$ , 则令  $m_{n+1} = m$ , 转到(2)。

观察  $K$  的计算式与  $K_n$  的计算式可知, 两者在形式上相差不多, 这似乎是说用  $K_n = K_{n-1} - K$  计算  $K_n$  与用  $K_n$  的原表达式计算  $K_n$  计算量差不多。事实上不然。在  $K_n$  的原表达式中求和是对所有与  $A$  相交的  $C_k$  来进行的, 在计算  $K$  的式子中, 求和是对所有与  $A$  不相交的  $C_k$  来进行的。在一个实际问题中, 不相交的项总要比相交的项少得多 (否则两个专家就太矛盾了), 所以说用  $K_n = K_{n-1} - K$  计算  $K_n$  还是有意义的。

在实际的决策问题中, 有时连状态集也不容易确定。在这种情况下, 利用专家咨询也可以同时确定出状态集来。参看第十一章 §2。

在对系统的生存状态的主观概率进行估计, 从而进一步作出决策的时候, 我们也可以首先利用第九章的一般方法, 在对各个专家进行咨询, 得到若干个基本可信度分配以后, 根据过程 MAXBEL 求到一个集合  $B$ 。求到  $B$  以后, 我们还可以对  $B$  中各单点的似真度归一化, 作为  $B$  中各状态生存的主观概率的估计。有了集合  $B$  及这些估计以后, 我们就可以将系统原来的状态集缩小到  $B$ , 然后针对新的状态集  $B$  作决策。显然, 这种方法是第九章一般方法的一个具体应用。详见下章 §3。

## §4 M 决策法

本节应用 Shafer 证据理论导出另一种决策分析的方法

——M决策法。

设状态集为  $S = \{x_1, \dots, x_q\}$ , 决策集为  $A = \{a_1, \dots, a_p\}$ , 报酬函数为  $r(a_i, x_j)$  ( $i = 1, \dots, p; j = 1, \dots, q$ )。

又设一批证据  $E$  在  $S$  上产生了一个基本可信度分配, 焦元为  $A_1, \dots, A_n$ , 可信数为  $m(A_1), \dots, m(A_n)$ 。令

$$\bar{r}(a_i, A_j) = \frac{1}{|A_j|} \sum_{x_k \in A_j} r(a_i, x_k)$$

$$(i = 1, \dots, p; j = 1, \dots, n)$$

选取目标为

$$V = \max_{a_i} \{E_r(a_i)\} = \max_{a_i} \{E(\bar{r}(a_i, A_j))\}$$

$$= \max_{a_i} \left\{ \sum_{j=1}^n \bar{r}(a_i, A_j) m(A_j) \right\}$$

若  $a^* \in A$  使  $E_r(a^*) = V$ , 则  $a^*$  即为所求的最优决策。

当  $r(a_i, x_j)$  取损失意义时, 在每个  $r(a_i, x_j)$  前加一负号上面的准则仍适用。

这种方法效果如何, 还需到实践中去检验。

**例10.4.1\*** 某公司下属三个工厂生产同一型号的电机, 为安排明年的生产计划, 对市场销售状况进行了分析, 可能的状况为差、一般、较好、好; 公司采取的方案为安排一个工厂, 二个工厂或三个工厂投产, 其报酬值见下页表10.1, 试确定最优方案。

该问题的状态集为  $S = \{x_1, x_2, x_3, x_4\}$ , 其中  $x_1, x_2, x_3, x_4$  分别表示销售状况差、一般、较好、好; 决策集  $A =$

---

• 该例取材于[21]。

表 10.1

报 酬 方 案	销 售 状 况	差	一般	较好	好
一个工厂投产		10	120	240	360
二个工厂投产		-80	200	360	480
三个工厂投产		-150	180	380	540

$\{a_1, a_2, a_3\}$ , 其中 $a_1, a_2, a_3$ 分别表示一个工厂投产, 二个工厂投产, 三个工厂投产。

设我们有一批证据 $E$ ,  $E$ 在框架 $S$ 上产生的基本可信度分配 $m$ 为

$$m(A) = \begin{cases} \frac{1}{3} & A = \{x_1, x_2, x_3\} \\ \frac{2}{3} & A = \{x_2, x_3, x_4\} \\ 0 & \text{其他} \end{cases}$$

这里 $m$ 的焦点为 $A_1 = \{x_1, x_2, x_3\}$ 和 $A_2 = \{x_2, x_3, x_4\}$ 。对 $A_1$ 和 $A_2$ 可计算 $\bar{r}(a_i, A_1)$ 和 $\bar{r}(a_i, A_2)$  ( $i = 1, 2, 3$ )

$$\bar{r}(a_1, A_1) = \frac{1}{3} (10 + 120 + 240) = \frac{370}{3}$$

$$\bar{r}(a_1, A_2) = \frac{1}{3} (120 + 240 + 360) = \frac{720}{3}$$

$$\bar{r}(a_2, A_1) = \frac{1}{3} (-80 + 200 + 360) = \frac{480}{3}$$

$$\bar{r}(a_1, A_2) = \frac{1}{3} (200 + 360 + 480) = \frac{1040}{3}$$

$$\bar{r}(a_2, A_1) = \frac{1}{3} (-150 + 180 + 380) = \frac{410}{3}$$

$$\bar{r}(a_2, A_2) = \frac{1}{3} (180 + 380 + 540) = \frac{1100}{3}$$

由此又可得针对焦元 $A_1$ 与 $A_2$ 的广义报酬值表(表10.2)。根据此表再计算 $E_r(a_i)$  ( $i=1, 2, 3$ )

表 10.2

广义报酬值 方案	焦元	$A$	$A_1$
$a_1$		$\frac{370}{3}$	$\frac{720}{3}$
$a_2$		$\frac{480}{3}$	$\frac{1040}{3}$
$a_3$		$\frac{410}{3}$	$\frac{1100}{3}$

$$E_r(a_1) = \pi(A_1) \cdot \bar{r}(a_1, A_1) + \pi(A_2) \cdot \bar{r}(a_1, A_2)$$

$$= \frac{1}{3} \times \frac{370}{3} + \frac{2}{3} \times \frac{720}{3} = \frac{1810}{9}$$

$$E_r(a_2) = \frac{1}{3} \times \frac{480}{3} + \frac{2}{3} \times \frac{1040}{3} = \frac{2560}{9}$$

$$E_r(a_3) = \frac{1}{3} \times \frac{410}{3} + \frac{2}{3} \times \frac{1100}{3} = \frac{2610}{9}$$

比较各 $E_r(a_i)$  ( $i=1, 2, 3$ ) 可得 $E_r(a_3)$ 为最大, 因此根据M决策法, 本例的最优方案为 $a_3$ , 即三个工厂同时投产该

种电机。

本例也可采用§2的专家咨询法先求得各状态生存的主观概率，继而求得最优方案。

由 $m$ 求 $pl(\{x_1\})$ ,  $pl(\{x_2\})$ ,  $pl(\{x_3\})$ ,  $pl(\{x_4\})$

$$pl(\{x_1\}) = \sum_{z_1 \in A} m(A) = \frac{1}{3}$$

$$pl(\{x_2\}) = \frac{1}{3} + \frac{2}{3} = 1$$

$$pl(\{x_3\}) = \frac{1}{3} + \frac{2}{3} = 1$$

$$pl(\{x_4\}) = \frac{2}{3}$$

归一化以后得各状态发生存在的主观概率的估计为

$$\begin{aligned} p(x_1) &= \frac{pl(\{x_1\})}{pl(\{x_1\}) + pl(\{x_2\}) + pl(\{x_3\}) + pl(\{x_4\})} \\ &= \frac{\frac{1}{3}}{\left(\frac{1}{3} + 1 + 1 + \frac{2}{3}\right)} = \frac{1}{9} \end{aligned}$$

$$p(x_2) = 1 / \left(\frac{1}{3} + 1 + 1 + \frac{2}{3}\right) = \frac{1}{3}$$

$$p(x_3) = 1 / \left(\frac{1}{3} + 1 + 1 + \frac{2}{3}\right) = \frac{1}{3}$$

$$p(x_4) = \frac{2}{3} / \left(\frac{1}{3} + 1 + 1 + \frac{2}{3}\right) = \frac{2}{9}$$

由此可求得各 $E_*(a_i)$  ( $i=1, 2, 3$ )

$$\begin{aligned}
E_r(a_1) &= r(a_1, x_1)p(x_1) + r(a_1, x_2)p(x_2) + r(a_1, x_3)p(x_3) + r(a_1, x_4)p(x_4) \\
&= 10 \times \frac{1}{9} + 120 \times \frac{1}{3} + 240 \times \frac{1}{3} + 360 \times \frac{2}{9} \\
&= \frac{1810}{9} \\
E_r(a_2) &= -80 \times \frac{1}{9} + 200 \times \frac{1}{3} + 360 \times \frac{1}{3} + 440 \times \frac{2}{9} \\
&= \frac{2560}{9} \\
E_r(a_3) &= -150 \times \frac{1}{9} + 180 \times \frac{1}{3} + 380 \times \frac{1}{3} + 540 \times \frac{2}{9} \\
&= \frac{2610}{9}
\end{aligned}$$

因此用这种方法所得的最优方案也为 $a_3$ ，即三个工厂同时投产。

例10.4.2 决策问题同例10.4.1。设我们有一批证据 $E$ ， $E$ 在框架 $S$ 上产生的基本可信度分配 $m$ ，焦元为 $A_1 = \{x_1, x_2\}$ 和 $A_2 = \{x_2, x_3, x_4\}$ ，基本可信数分别为 $m(A_1) = \frac{4}{5}$ ， $m(A_2) = \frac{1}{5}$ 。

则用 $M$ 决策法可求得（过程从略）

$$E_r(a_1) = \frac{600}{6} \quad E_r(a_2) = \frac{704}{6} \quad E_r(a_3) = \frac{512}{6}$$



用专家咨询法可求得（过程从略）

$$E_r(a_1) = \frac{1240}{11} \quad E_r(a_2) = \frac{1520}{11} \quad E_r(a_3) = \frac{1220}{11}$$

可见无论采用哪种方法其结果都是最优方案为  $a_2$ 。

在本例中用两种方法所得到的期望报酬值是不一样的，这说明例10.4.1中期望报酬值相同的情况是偶然的。但是，在这两个例子中两种方法的决策结果却都是相同的。这一点提醒我们去猜测，对于同一个决策问题，用M决策法和用专家咨询法做决策，其结果是否总是一致的。不过这个结论只有经过理论证明以后才能真正成立。

设针对状态集  $S$ ，我们没有一点证据。在这种情况下，我们所得到的基本可信度分配为  $m(S) = 1$ ， $m(\text{其他}) = 0$ 。也即在没有任何证据可利用的情况下，焦点只有  $S$  本身。因此利用M决策法求最优方案时， $f(a_i, S) (i = 1, \dots, p)$  的计算就相当于对特定方案  $a_i$ ，求在各种可能状态下报酬值的平均值。而最后的准则，由于  $m(S) = 1$ ，所以求  $E_r(a_i) (i = 1, \dots, p)$  的最大值就相当于求  $f(a_i, S) (i = 1, \dots, p)$  的最大值。因此利用M决策法求最优方案就相当于在不确定型决策中利用等可能性准则求最优方案。所以，在没有任何证据可利用的情况下，M决策法退化为等可能法，或反过来说M决策法是等可能法有证据时的推广。

另外，上面讨论的决策分析的方法（包括传统的决策分析的方法）是边估计主观概率边进行决策。事实上，我们也可以采用先预测系统的状态，然后再进行决策的方法。详见第十一章§3。

## 第十一章 预 测

本章是第九章利用Shafer证据理论作决策的一般方法的一个典型应用。

### §1 预测及常用方法

预测就是对未来作出估计，更详细地说，预测是对某研究对象的未来状态进行预计和推测。<sup>[22]</sup>

对未来作出估计不能靠求神拜佛，不能靠问卜决疑，也不能靠某个人的主观想象。换句话说，预测要方法得当，有根有据。

具有关资料的统计。预测方法多达二百余种。<sup>[22]</sup>但是，根据它们所依据的原理，大体上可以分为依据惯性原理的方法，依据类推原理的方法和依据相关原理的方法。

客观事物的发展过程常常表现出它的延续性，即过去、现在怎么样，将来也不会有多大的变化。因此根据过去和现在的情况就可以推断将来的趋势。此即惯性原理。根据惯性原理作预测，关键点是我们必须对过去和现在的情况有充分的了解，即必须拥有大量的历史资料和记载。只有这样，才

能作出比较理想的预测。可以说,所有基于概率理论的方法,如时间序列预测,马尔科夫预测,增长曲线预测等都属于这样的类型。

类推原理基于特性相近的客观事物其变化有相似之处这样一个基本原则。所以,使用类推原理进行预测,首要的是寻找特性相近的事物,然后根据两事物的联系,寻找已知事物的变化特征,从而推断未知事物的未来状态。例如,根据甲国的能源消耗量类推乙国,根据军用飞机的最大飞行速度预测民航客机的最大飞行速度。

任何事物的变化都不是孤立的,而是在与其他事物的相互影响下发展的,相互影响常常表现为因果关系。如果两个事物互为因果,那么我们就可以由原因事物推出结果事物。此即相关原理。根据相关原理,我们要预测某事物未来状态,我们必须弄清它所处的环境,然后根据环境的变化情况,预测该事物的变化。所有因果型预测法,如回归分析法都基于这样的原理。

值得说明的是,有许多方法很难归到上面所说的任何一个类型。原因是很难说出它依据什么样的原理(因为它所依据的原理不止一个)。大多数定性预测都是这样。这里,我们要特别指出的是特尔菲法。<sup>[22][29]</sup>

特尔菲法是由美国兰德公司于1964年发明并首先用于技术预测的,不过后来证明这种方法用于其他领域也是相当成功的。尤其是对于复杂的社会预测,这种方法的优越性更加明显。

应用特尔菲法进行预测有几个关键问题要考虑。

1. 专家的选择。特尔菲法依靠专家作预测,所以对专家

的选择要特别慎重。要选择那些对本专业有很深的了解，有预见能力的专家。

2.调查表的设计。将来专家的意见就要通过调查表来反映，所以说，调查表应该设计的有利于专家作答，并能全面反映专家意见。

3.专家意见的处理。对专家意见应作出适当的处理，即对调查表应作出适当的处理。过去，对专家意见通常作出一些统计处理，即用统计的办法综合专家意见。

当然，设立预测机构，向专家介绍特尔菲法，对专家支付适当的报酬等问题也必须引起足够的重视，限于篇幅，不再详述，可参看[22][23]。

## ✓ §2 信度预测法

本节应用Shafer证据理论给出另一种预测的方法。这种方法以第九章利用Shafer信度理论作决策的一般方法为基础，类似于上章主观概率估计的专家咨询法。

设我们的预测问题是：要求预测出某个领域未来某个时刻发生的事件，并给出该事件发生的可能性。

对于该问题，同样可以采用特尔菲法进行预测。

应用Shafer的证据理论可以设计出一种新的特尔菲咨询方式，而且可以对咨询的结果作出比较容易的处理。

设请了 $p$ 个专家，对每一个专家都要请他回答这样的问题：请你指出几个未来可能发生的事件的集合，并且指出你有多大把握作出这种判断。

对于一个该领域的专家，要他回答这样的问题是不困难

的。任何一个专家根据他自己的经验和知识，经过慎重的、仔细的思考，总能说出这样的话：“可能发生的事件80%跑不出什么、什么和什么。”说这样的话对专家来讲是非常自然的，常见的。于是我们就可以根据每位专家说出的这些话作出综合性的预测。所以说，我们这种方法不需设计很复杂的调查表，咨询方式也是非常简单的，甚至不需要组织专门的形式，在调查人员与专家的对话中即可解决。

这一步工作做完以后，将每个专家指出的所有事件的集合合并起来作为我们的识别框架 $\Theta$ 。

设第 $i$ 个专家指出的未来可能发生的事件的集合为 $A_{i1}, A_{i2}, \dots, A_{ik}$ ，所对应的把握为 $r'_{i1}, r'_{i2}, \dots, r'_{ik}$ （实际情况可能没有这么复杂，一个专家至多指出一两个可能发生的事件的集合）。令

$$r_{ij} = \frac{r'_{ij}}{\sum_{i=1}^k r'_{ij}} \quad (j = 1, \dots, k)$$

则第 $i$ 个专家所对应的证据 $E_i$ 在框架 $\Theta$ 上产生的基本可信度分配 $m_i$ ，满足

$$\begin{aligned} m_i(A_{ij}) &= r_{ij} \quad (j = 1, \dots, k) \\ m_i(B) &= 0 \quad B \neq A_{ij} \quad (j = 1, \dots, k) \end{aligned}$$

根据第九章，由Dempster合成法则可求出这 $p$ 个基本可信度分配的直和 $m$ ，然后由 $m$ 求出所对应的信度函数 $Bel$ ，由 $Bel$ 依据过程MAXBEL求得集合 $B$ ，则 $B$ 即为这次预测的结果，即未来可能发生的事件的集合。

$$\forall \theta \in B, \text{ 用 } pl(\{\theta\}) = \sum_{A \in \mathcal{A}} m(A) \text{ 计算单点似真度 } pl(\{\theta\}),$$

归一化以后即可作为各事件发生的可能性。

如果对以上预测还不满意，那么我们可以进行第二轮的咨询。此时令 $\Theta = B$ ，提问的方式可改为：“请指出在 $\Theta$ 中哪些子集合包含未来最可能发生的事件，并指出你有多大把握作出这种判断。”

咨询结束以后，同样得到若干基本可信度分配，用第九章的方法同样可求得一个集合 $B$ ，该集合 $B$ 肯定是上一步得到的集合 $B$ 的子集，即我们的预测的结果集缩小了。

以后对于上一步的结果集还可进行第三轮、第四轮……的咨询，直至作出比较满意的预测为止。

✓ 由上可知，这种方法是逐步缩小预测范围的方法。

应用这种方法作预测，同样需要考虑我们上节提到的应用特尔菲法预测时应注意的几个关键问题，比如，专家的选择，提问方式的选择（上面说的提问方式是基本原则，具体的提问方式可以设计得较好些，以便更进一步有利于专家作答，甚至可考虑设计启发式的提问方式）等。关于其他问题，比如说机构设置，对专家支付适当的报酬，应用这种方法预测时同样存在，在此不再赘述。

由以上分析可以看出，信度预测法是特尔菲预测法的推广。在提问方式的设计上，专家意见的描述上以及专家意见的处理上都与特尔菲预测法有本质的区别。

信度预测法除适用于以上问题的预测外，还适用于其他同类型的预测问题。如系统所处的状态的预测用信度预测法来进行也是非常合适的。

### §3 另一种决策分析的方法

设系统的状态集为 $\Theta = \{x_1, \dots, x_n\}$ , 该状态集是系统可能生存的状态。但是一个系统所处的状态仅仅是一种, 如果我们能确定出这种状态, 那么我们的决策就变成了确定型的决策; 如果我们不能确定出这种状态, 那么我们的决策就是不确定型决策。有时, 虽然我们的决策是不确定型决策, 但是, 如果我们能想方设法排除一些不太可能出现的状态, 然后集中考虑若干最有可能出现的状态, 针对这些状态作决策, 那么我们决策的不确定性也会有所减少。

将上节的预测方法应用于状态集 $\Theta = \{x_1, \dots, x_n\}$ 即可起到缩小状态集的目的。

设对状态集 $\Theta = \{x_1, \dots, x_n\}$ , 应用上节的方法作预测, 经过若干轮咨询与处理以后所得的最后结果为集合 $A = \{x_{i_1}, \dots, x_{i_n}\}$ , 可能性为 $P(x_{i_j}) = p_{i_j} (j=1, \dots, n)$ 。显然, 一般情况下集合 $A = \{x_{i_1}, \dots, x_{i_n}\}$ 是状态集 $\Theta = \{x_1, \dots, x_n\}$ 的真子集,  $A$ 中的状态为 $\Theta$ 中最有可能出现的状态, 因此, 可针对集合 $A$ 作决策。针对集合 $A$ 作决策显然要比针对 $\Theta$ 作决策不确定性小的多。

将 $A$ 作为新的状态集,  $P(x_{i_j}) = p_{i_j}$ 作为其主观概率, 在报酬矩阵中去掉与 $A$ 中状态无关的列, 则可依照决策分析中有关的方法作出决策。

值得说明的是, 经过预测,  $A$ 中元素与 $\Theta$ 中元素相比要少得多 (有时仅剩1个), 而针对很少的几种状态作决策, 显然要比针对许多种状态作决策容易的多。

上章讨论的决策分析的方法是边估计系统状态生存的主观概率，边进行决策，本章讨论的方法是先预测状态，然后进行决策。可见本章的方法是将一种决策转化成另一种决策的方法。



## 第十二章 证据理论的另一种类型的应用

对于Shafer证据理论，我们有一个隐含的假设，就是在框架 $\Theta$ 中存在且仅存在一种可能性是该判决问题的答案，即在 $\Theta$ 中存在着唯一的真值。如，要鉴别一件出土文物的年代，我们的框架 $\Theta$ 取为所有可能的年代的集合，但在框架 $\Theta$ 中只存在着唯一的一个年代是该出土文物的年代（一件文物不可能既是元朝的，又是清朝的），所以对于框架 $\Theta$ ，我们可以应用Shafer证据理论进行处理。

本书第九章利用Shafer证据理论作决策的一般方法以Shafer证据理论为基础，所以应用这个方法作决策时也必须遵循这样一个假设。比如第十章决策分析中主观概率的估计。我们的问题是找出各种可能的状态发生的概率。此处，我们取框架 $\Theta$ 为状态集 $S$ 。在框架 $\Theta$ 中存在着唯一的真值，即系统未来所处的状态只能是一种，所以对于该问题，我们可以利用第九章的方法来解答。再比如，第十一章预测问题的讨论。我们选取框架 $\Theta$ 为所有未来某个时刻可能发生的事件，而未来某个时刻发生的事件只能是一种，即在 $\Theta$ 中只有一个真值，

所以对于预测问题，我们同样可以应用第九章的方法来解答。

但是也有许多其他类型的问题，初看起来不满足这样一个隐含的假设，因而不能直接应用Shafer证据理论作决策。但是经过问题变形以后，仍然可以满足该假设，因而仍然可以利用Shafer证据理论作决策。这样的问题很多，本章仅以指标体系的确定（§1），领导班子的选取（§2），医疗诊断（§3）为例来说明这一类问题的解决办法。

## §1 指标体系的确定

在[24]中，伊·彼·苏斯洛夫讨论了统计指标的概念，作用、职能、哲学性、形成方法以及种类和体系等问题。一项统计指标是说明社会现象（单个和总体的）某种特征的数字。说明个体现象的是个体指标，说明总体现象的是综合指标。

统计指标能表明社会生活的一切领域的特征。如果是表明经济状况的，就叫做经济指标，如果是涉及政治生活的就叫做政治指标。本文所说的统计指标如果不声明的话，是就一般的统计指标而言的，当然对于特殊的经济指标、社会指标等也是适用的。

一个统计指标体系是针对一定的对象和一定的研究目的而言的，是一个反映了各种标识和现象现实存在的相互联系的指标的辩证统一体。

设针对一个社会现象，反映该现象的各个属性的所有可能的指标（目前可认识到的）为 $F = \{f_1, f_2, \dots, f_n\}$ 。现

在的问题是：为了研究某个问题，采取什么样的指标体系比较合适，即如何从集合中选出一个子集来，使得对于某个问题研究起来最好（最能反映该现象）。

假设我们就针对 $F$ 应用Shafer证据理论确定指标体系，那么我们必须假设：在描述一个对象的指标集合 $F$ 中存在一个唯一能完全描述该对象的指标 $f$ ，只是由于我们的证据不足，不能确定出来（完全是一种假设），于是在一定的证据下，我们只好来寻求最有可能包含那样一个指标 $f$ ，而且研究起来最方便的子集合 $A$ 代替 $f$ 来描述该对象，在这个意义上，子集 $A$ 就被理解成了描述该对象的一个较合适的指标体系。

但是，这种假设和这种理解看起来是那样的牵强附会，以至于将来我们利用这个假设得出结果以后经济学家不承认这样一个结果。

现在，指标体系是 $F$ 的一个子集这一点似乎是无可非议的。因为，假设又有某个专家构造出一种指标来，而且要尝试将该指标用于指标体系，那么我们可以将该指标加到集合 $F$ 中。事实上，一开始我们就假设 $F$ 包括了所有目前已认识到的反映该现象的指标，所以在目前的认识水平上， $F$ 不存在遗漏指标的问题。所以，认为指标体系是 $F$ 的一个子集是相当自然的。

指标体系是 $F$ 的一个子集，那么它就是 $\Theta = 2^F$ （ $F$ 的幂集）中的一个元素，所以说指标体系是 $\Theta$ 中的一个元素。进一步，我们假设指标体系是 $\Theta$ 中的唯一一个元素。从而针对 $\Theta$ ，我们可以应用第九章的方法求指标体系。

此处，也许有人对只存在唯一的指标体系有看法。他们的理由是，有许多统计学家或经济学家都提出了各自认为是

最好的指标体系。事实上,针对某个特定问题可能会有许多个指标体系,有的人采取这一个,有的人采取那一个。但是,这不能说明指标体系不唯一。相反,存在这么多的被各人称作指标体系的指标体系恰恰说明,大家都在力争寻找这一个唯一的指标体系。

在 $\Theta$ 中,有一些集合,比如说空集、单点集以及整个 $F$ ,都不可能作为指标体系,因此可以从 $\Theta$ 中删除掉。另外,其他一些明显不能作为指标体系的指标集也可从 $\Theta$ 中删除掉。假设删除以后的集仍然叫做 $\Theta$ 。

$\Theta$ 确定以后,我们就可以利用第九章的方法以及预测中的信度预测法找指标体系了。

与前几章讨论相同。设有 $p$ 个专家,现在请每一个专家就 $\Theta$ 而言选定一个或几个他认为有可能包括指标体系的子集,而且还要请他指出他有多大把握认为这几个子集中的某一个可以包括指标体系。因为我们请的都是专家,所以这些问题他们是能够回答的。不过,对于提问方式和对专家的待遇等问题要精心设计,认真对待。总之要想方设法使专家愿意认真地思考这些问题,并认真地作答(这些问题与特尔菲法预测<sup>[22]</sup>中所考虑的相类似)。这样,可以假设,每个专家都是根据他的知识、他以往的经验以及对该问题、该对象的充分的分析和研究作出这些回答的。所以我们可以将这些答案看成是根据充分的证据所得到的。而且假设如果我们是个专家的话,也即我们与那位专家所具有的知识、经验等相同的话,也会得出与他相同的答案,所以咨询结束以后,我们可以认为我们自己根据不同的证据得到了一些不同的结果。

设第 $i$ 个专家指出的集合为 $A_{i_1}, A_{i_2}, \dots, A_{i_k}$ , 他的把

握分别为 $r'_{i_1}, r'_{i_2}, \dots, r'_{i_k}$ , 令

$$r_{ij} = \frac{r'_{ij}}{\sum_{i=1}^k r'_{ij}} \quad (j=1, \dots, k)$$

(对 $r'_{i_1}, r'_{i_2}, \dots, r'_{i_k}$ 归一化)

则我们可以认为第 $i$ 个专家所对应的那批证据 $E_i$ 在 $\Theta$ 上产生了一个基本可信度分配 $m_i: 2^\Theta \rightarrow [0, 1]$ 。其中

$$m_i(A_{i_j}) = r_{ij} \quad (j=1, \dots, k)$$

$$m_i(B) = 0 \quad (B \neq A_{i_1}, \dots, A_{i_k})$$

得到这 $p$ 个 $m_i$ 以后, 将这 $p$ 个 $m_i$ 按Dempster合成法则(可用(3.2.1)式也可以用第十章§3的递推形式)进行合成可得基本可信度分配 $m$ , 根据 $m$ , 求得对应的 $Bel$ , 然后利用第九章中的过程MAXBEL求得 $\Theta$ 的一个子集 $B$ 。

得到集合 $B$ 以后, 可以像上章处理预测问题那样,  $\forall A \in B$ ,

用 $pI(\{A\}) = \sum_{A \in C} m(C)$ 计算单点似真度, 然后求出使单点似

真度最大的 $A$ 即可作为我们的指标体系。

如果 $B$ 中单点似真度最大的 $A$ 不唯一, 则可取 $\Theta = B$ , 重复上述过程即进行第二轮咨询, 第三轮咨询……, 直到求到满意的指标体系为止。

另外, 如果一开始 $F$ 我们不能确定的话, 通过咨询可以同时确定出 $F$ 来。提问方式可以变为: “请你指出一些有可能作为指标体系的指标的集合来, 而且也要指出你有多大把握认为这样的集合可以作为指标体系。”最后将每一个专家所说的所有可能作为指标体系的集合并起来作为我们的 $F$ 。不过, 与上章不同, 以下的处理方法要作些调整。

设第 $i$ 个专家指出的可作为指标体系的指标的集合分别为 $Y_{i1}, Y_{i2}, \dots, Y_{in}$ , 所对应的把握分别为 $r_{i1}, r_{i2}, \dots, r_{in}$ , 则我们可以认为第 $i$ 个专家所指出的也就是在框架 $\Theta = 2^U$ 中指标体系不会跑出集合 $Y_i = \{Y_{i1}, Y_{i2}, \dots, Y_{in}\}$ , 而且他的把握是 $r_i = \max_{j \in \{1, \dots, n\}} r_{ij}$ 。据此, 我们可以得到第 $i$ 个专家所对应的

证据 $E_i$ 在框架 $\Theta$ 上产生的基本可信度分配为

$$\begin{cases} m_i(Y_i) = r_i \\ m_i(\Theta) = 1 - r_i \\ m_i(\text{其他}) = 0 \end{cases}$$

显然, 该基本可信度分配所对应的信度函数为简单支持函数。因此利用Dempster合成法则合成这 $p$ 个简单支持函数以后所得到的直和 $S$ 是一个可分离支持函数。

设针对可分离支持函数 $S$ 利用第九章的过程MAXBEL所得到的集合为 $B$ , 则与上相同, 我们可以 $\forall A \in B$ 求单点似真度 $pl(\{A\})$ 。求出所有单点似真度以后, 我们可以选取单点似真度最大的集 $A$ 作为我们的指标体系。如果单点似真度最大的集 $A$ 不唯一, 那么同样必须再向专家咨询, 直到找到我们认为合适的指标体系为止。

用该法可以综合各专家的意见, 是我们第九章集体决策法的另一种体现。

## §2 领导班子的选取

如何选取领导班子, 在企业或其他组织中的管理体制中占有比较重要的地位。

领导班子的选取通常依靠投票的办法。投票可以分为有候选人的投票和无候选人的投票两种。但不管哪一种都是根据被选者的得票多少决定谁被选上，谁没被选上或谁被选为第一把手，谁被选为第二、三……把手。

这种选取特别适合于从一个群体中选取1个或几个优秀分子的情况。因为根据得票者的得票多少决定谁被选上，谁没被选上可以反映大多数群众的看法。但是，这种方法却不一定适合于领导班子的选取，原因是，领导班子的选取不能单单考虑某个被选的领导是否受群众欢迎或他自己的工作能力是否强这样一个因素。领导班子是一个集体，领导班子工作的好坏取决于领导班子中的所有成员互相配合、协调地工作的能力，而不取决于领导班子中的一个一个的成员单独工作的能力。因此，依据得票多少一个一个地选举产生领导班子成员的选举办法，由于没有考虑领导班子的集体性，因而是不妥当的。

现在，也有许多组织，领导班子是靠下列方法产生的。首先由群众选举产生第一把手，然后由第一把手去组阁。这样做的好处是领导班子比较团结，领导与领导之间相互配合的意识较强。但是，由于第二把手以下的领导都是第一把手推荐的，因此容易出现任人为亲的局面，容易出现第一把手的独裁统治，长期下去，领导班子的集体性就会被破坏。因而这种方法也是不可取的。

这就是说，在领导班子选取的时候，必须考虑组织中的哪些成员组合起来能够最好地领导整个组织的问题。这里，没必要规定需要多少个成员去组合（即不要规定需要几个经理、几个委员），相反，应该考虑既能够很好地领导整个组

织，班子中的人员又最少的那样的领导班子。三个人能干的很好的事情，为什么非要四个、五个甚至更多的人干呢？

经常听到好多人议论：“要是让张三，李四再加上王五领导我们这个科，那我们这个科早就上去了。”“要是让A去配合我们经理工作，那我们经理肯定能干好。”这就是说，实际上群众考虑的也是由哪些人组合起来才能更好地领导他们这个组织的问题。

因此选领导班子的时候，我们可以让群众这样来投票，“你认为本组织中哪些人组合起来能够最好地领导该组织，请写出这些人的名单，并指出你有多大把握作出这样的判断。”

这样投票不需要预先指定候选人，也不需要考虑本组织中谁最有领导能力（上面说过，最有领导能力的人不一定适合进入领导班子）。只要指出，整个领导班子应该由什么人来担任即可！

值得说明的是，有的人可能不只说出一个合适的领导班子。这没关系，在这种情况下只要明确指出，你有多大把握认为它们合适即可！

应用第九章的方法，我们可以对这些选票作出处理。

设该组织所有成员的集合为  $F = \{x_1, \dots, x_p\}$ ，我们要选的领导班子是  $F$  的一个子集或是  $\Theta = 2^F$  中的一个点。在  $\Theta$  中去掉  $\emptyset$ 、 $\Theta$  以及那些明显不能作为领导班子的子集，设所得到的集合仍用  $\Theta$  来表示，就取  $\Theta$  作为我们的识别框架。

在作了这种假设以后，由每张选票就可以得到  $\Theta$  上的一个基本可信度分配。

设第  $i$  张选票指出的领导班子的集合为  $A_{i1}, \dots, A_{ik}$ ，所对应的把握为  $r_{i1}, \dots, r_{ik}$ ，则我们可以认为第  $i$  张选票指出



了在框架 $\Theta$ 中领导班子的集合跑不出集合 $A_i = \{A_{i_1}, \dots, A_{i_k}\}$ , 而且所对应的把握为

$$r_i = \max_{j \in \{1, \dots, k\}} r_{i_j}$$

由此我们可以得到第 $i$ 张选票所对应的基本可信度分配 $m_i$ 满足

$$\begin{cases} m_i(A_i) = r_i \\ m_i(\Theta) = 1 - r_i \\ m_i(\text{其他}) = 0 \end{cases}$$

由Dempster合成法则求出这 $p$ 个基本可信度分配的直和 $m$ , 根据 $m$ 求出所对应的信度函数Bel (易见为可分离支持函数), 对信度函数Bel, 根据过程MAXBEL求得一个集合 $B$ 。

$\forall A \in B$ , 用 $pl(\{A\}) = \sum_{A \in C} m(C)$ 计算 $B$ 中各个集合的似真度, 设似真度最大的集合为 $D$ , 则 $D$ 即为这次投票选举出的领导班子 (如果似真度最大的集不唯一, 则可由群众举手表决确定出唯一一个集)。

尽管对于这样选举出的领导班子也会有人提出一些不赞成的地方, 但是它综合了全体群众的意见, 是群众选举的结果, 符合大多数群众的看法。

### §3 医疗诊断

当一个人患了某种疾病时, 必定具有某种反应。例如, 一个人如果患了感冒, 则一般情况下要发烧、头痛、流鼻涕, 这种反应称为症状。一定的疾病有一定的症状, 所以可以根据一定的症状判断出某人患什么疾病。但是, 一种症状有时

可以被多种疾病所具有。比如，发烧可以由感冒引起，也可以由肺炎引起。所以出现某种症状时，往往不能立即诊断出病人患什么病，而是得到一个病人可能患病的集合。要想进一步诊断出病人患什么病，只能再进一步观察症状。

可是，有时症状多并不能给确诊带来多大好处。原因是，在这么多的症状中，难免出现有些症状彼此干扰，互相对立，有些症状彼此联系，互相诱发的情况。因此，症状的选取必须抓主要矛盾。可是，哪些症状是主要的，哪些症状又是次要的呢？这一点是很难确定的。再加上一个人所患疾病经常出现两种以上的情况，这就给诊断带来了更大的困难。在这种情况下，为了更好地诊断，医学上通常采用专家会诊的方式。利用第九章的决策方法，我们可以给这种专家的会诊以一定程度上的数量描述。

✓ 设病人所可能患病的集合为 $F$ ，取 $\Theta = 2^F$ 为我们的识别框架，同上所述，各个专家的会诊都可以在 $\Theta$ 上产生一个基本可信度分配，根据Dempster合成法则求它们的直和，再求直和所对应的信度函数Bel，利用过程MAXBEL求得 $\Theta$ 的一个子集 $B$ ，求出 $B$ 中单点似真度最大的集合，记为 $A$ ，则 $A$ 即可作为病人患病的集合。如果似真度最大的集合不唯一，则可进一步会诊，其方法同上（只要令 $\Theta = B$ 即可！）。

根据上面这种数量描述，我们可以模拟专家的会诊，制成各种医疗诊断专家系统，这对大大丰富电子计算机在医疗诊断中的应用必将起到非常积极的作用。

## 第十三章 人工智能与专家系统

本章探讨证据理论在人工智能以及专家系统中的应用。

其中§1, §2, §3探讨了证据理论在推理中的应用; §4将证据理论应用于模式识别问题, 得到了模式识别问题的另一种处理方法; §5探讨了专家系统中各种不确定性的存在, 综述了现有专家系统对这些不确定性的处理, 最后论述了Shafer 证据理论对不确定性的处理, 提出了设计基于信度函数而不是基于产生式规则的专家系统的思想。

### §1 软推理

Shafer在[1]中一开始的形式化体系中, 使得命题演算转化成了比较直观的集合运算。在本节中我们假设已经有了一个识别框架 $\Theta$ , 而且所有我们感兴趣的命题都可以用 $\Theta$ 的一个子集来表示。另外, 我们也不区分命题与集合, 比如, 若 $B \subset A$ 则我们就说 $B$ 蕴含 $A$ 。

**定义13.1.1** 如果 $B \subset A$ 很松散地成立,  $B$ 真, 则直观上我们可以推出 $A$ 大概真。

如果 $B \subset A$ 很松散地成立,  $A$ 真, 则直观上我们可以推出

$B$ 比较可靠。

如果 $A \cap B \neq \phi$ 成立,  $A$ 真, 则直观上我们可以推出  $B$  似乎真。

这些近似推理称为软推理。特别是第三种称为相容推理。

说明几点:

(1)  $B \rightarrow A (B \subset A)$ 很松散地成立是指蕴含本身的程度很小, 即蕴含关系很松散, 很勉强 (也许是经过许多步推理以后才得到的) 或即使蕴含关系很强, 但是由于证据不足, 我们对它的信任程度很小 (一种主观看法)。

(2) 这里的  $A, B$  的真假并非客观上的真假, 而是我们的主观看法, 所以用对它为真的信度来描述, 如果信度为1, 说明它在我们心目中被认为是真的, 小于1说明有点真但不太真。

(3) 相容推理是一种更加软的推理。 $A \cap B \neq \phi$ 意味着  $A$  与  $B$  具有一定的联系。所以当其中一个有点真的时候, 另一个也必定表现出某种程度的似真性。 $A \cap B$ 大, 说明这种联系比较紧密, 因而表现出的似真性就大;  $A \cap B$ 小, 说明这种联系比较松散, 因而表现出的似真性就小。

由此可以看出, 这种软推理更加接近人类进行推理的真正模式, 更加符合人的习惯。

应用Shafer的信度理论有可能使这些软推理得到数量上的描述。

设原先在 $\Theta$ 上已经有一个信度函数 $\text{Bel}_1$ , 现在又来了一批新证据, 这批证据使我们确信 $B$ 真, 则这批新证据又在框架 $\Theta$ 上产生了一个信度函数 $\text{Bel}_2$ 。

$$\text{Bel}_2(A) = \begin{cases} 1 & B \subset A \\ 0 & B \not\subset A \end{cases}$$

则由(3.5.2)式知: 当且仅当  $\text{Bel}_1(\bar{B}) < 1$  时

$$\begin{aligned} \text{Bel}_1(A|B) &= (\text{Bel}_1 \oplus \text{Bel}_2)(A) \\ &= \frac{\text{Bel}_1(A \cup \bar{B}) - \text{Bel}_1(\bar{B})}{1 - \text{Bel}_1(\bar{B})} \quad (13.1.1) \end{aligned}$$

此即 Dempster 的条件规则。

Dempster 的条件规则还可以用似真度函数和众信度函数来表示。

第二批证据产生的似真度函数为

$$\begin{aligned} \text{pl}_2(A) &= 1 - \text{Bel}_2(\bar{A}) = 1 - \begin{cases} 1 & B \subset \bar{A} \\ 0 & B \not\subset \bar{A} \end{cases} \\ &= \begin{cases} 1 & A \cap B \neq \phi \\ 0 & A \cap B = \phi \end{cases} \end{aligned}$$

当且仅当  $\text{pl}_1(B) = 1 - \text{Bel}_1(\bar{B}) > 0$ , 即  $\text{pl}_1(B) \neq 0$  时

$$\text{pl}_1(A|B) = (\text{pl}_1 \oplus \text{pl}_2)(A) = \frac{\text{pl}_1(A \cap B)}{\text{pl}_1(B)} \quad (13.1.2)$$

第二批证据产生的众信度函数为

$$\begin{aligned} Q_2(A) &= \sum_{C \subset A} (-1)^{|C|} \text{Bel}_2(\bar{C}) \\ &= \sum_{C \subset A} (-1)^{|C|} \begin{cases} 1 & \text{如果 } B \subset \bar{C} \\ 0 & \text{如果 } B \not\subset \bar{C} \end{cases} \\ &= \sum_{\substack{C \subset A \\ C \cap B = \emptyset}} (-1)^{|C|} \\ &= \sum_{C \subset A \cap \bar{B}} (-1)^{|C|} \end{aligned}$$

$$\begin{aligned}
&= \begin{cases} 1 & A \cap \bar{B} = \Phi \\ 0 & A \cap \bar{B} \neq \Phi \end{cases} \quad (\text{见附录 §6}) \\
&= \begin{cases} 1 & B \supset A \\ 0 & B \not\supset A \end{cases}
\end{aligned}$$

(此结论也可直接由  $Q$  的定义得出)

当且仅当  $\sum_{C \in B} (-1)^{|C|} Q_1(C) < 1$  时

$$\begin{aligned}
Q_1(A|B) &= (Q_1 \oplus Q_2)(A) = K Q_1(A) Q_2(A) \\
&= \begin{cases} K Q_1(A) & A \subset B \\ 0 & A \not\subset B \end{cases} \quad (13.1.3)
\end{aligned}$$

$$\begin{aligned}
\text{其中} \quad K &= \left( \sum_{\substack{C \in B \\ C \neq \Phi}} (-1)^{|C|+1} Q_1(C) Q_2(C) \right)^{-1} \\
&= \left( \sum_{\substack{C \in B \\ C \neq \Phi}} (-1)^{|C|+1} Q_1(C) \right)^{-1} \\
&= \left( 1 - \sum_{C \in B} (-1)^{|C|} Q_1(C) \right)^{-1}
\end{aligned}$$

Dempster 的条件规则说明了一开始我们对各个命题(集合)有一信度, 现在一批新的证据使我们确信  $B$  真时, 我们对各个集合的信度、众信度、似真度的变化。所以根据 Dempster 条件规则, 我们有可能对上面定义的那些软推理以进一步的讨论。

如果  $B \subset A$  很松散地成立,  $B$  真, 则由 (13.1.1) 式知

$$\text{Bel}_1(A|B) = \frac{\text{Bel}_1(A \cup \bar{B}) - \text{Bel}_1(\bar{B})}{1 - \text{Bel}_1(\bar{B})}$$

该式说明:  $A$  (结论) 变化后的信度大小随着我们对  $B$  蕴含  $A$  的信度即  $\text{Bel}(\bar{B} \cup A)$  的增加而增加, 随着前件  $B$  的非成

立的信度的增加而减小（因为它是  $\text{Bel}_1(\bar{B})$  的递减函数），即  $B$  蕴含  $A$  越可信则由  $B$  推出的  $A$  越可信，前件  $B$  越不成立则推出的  $A$  越不可信。

由 (13.1.2) 式知

$$\text{pl}_1(A | B) = \frac{\text{pl}_1(A \cap B)}{\text{pl}_1(B)} = \frac{\text{pl}_1(B)}{\text{pl}_1(B)} = 1$$

说明尽管  $B \supset A$  是松散的，但是  $B$  真则  $A$  也似乎真。

如果  $A \supset B$  很松散地成立， $B$  真， $A$  怎么样呢？

由 (13.1.2) 式知

$$\text{pl}_1(A | B) = \frac{\text{pl}_1(A \cap B)}{\text{pl}_1(B)} = \frac{\text{pl}_1(A)}{\text{pl}_1(B)}$$

该式说明： $A$  变化后的似真度与开始我们对  $A$  的似真度成正比，与我们认为  $B$  似乎真的程度成反比。即我们开始认为  $B$  越真，则证实  $B$  以后对  $A$  的似真度影响越小（增加越少）。

由 (13.1.3) 式知

$$\text{Q}_1(A | B) = \text{KQ}_1(A)$$

该式说明  $A$  变化后的众信度与我们开始对  $A$  的众信度成正比。

第一种推理不能用 (13.1.3) 式表示，第二种推理不能用 (13.1.1) 式表示，这除了从数学形式上不允许以外，这两种推理的信息类型也不允许这样（见[13]）。

如果  $A \cap B \neq \emptyset$  成立， $B$  真，则由 (13.1.2) 式知

$$\text{pl}_1(A | B) = \frac{\text{pl}_1(A \cap B)}{\text{pl}_1(B)}$$

若  $A \cap B$  越大，说明  $A$  与  $B$  越相像，则  $\text{pl}_1(A \cap B)$  越大，从

而 $pl_1(A|B)$ 越大；若 $B$ 的似真度越大，则 $pl_1(A|B)$ 越小，即 $B$ 原来看起来越真，则验证 $B$ 成立以后， $A$ 的似真度增加越少。

由上分析可见，应用Shafer证据理论来解释这些推理基本上符合人们的直观想象。

这些推理的不确定性在专家系统中类似于规则的不确定性。因此，这些推理的研究对于专家系统中不确定性的处理具有重要的意义。

## §2 用证据理论解决归纳问题

什么是归纳问题？简单地讲，归纳就是由特殊推出一一般。详细地说，归纳是由某些特殊事例所具有的特性推测所有的一般的事例所具有的特性。比如说，由甲种鸟会飞，乙种鸟也会飞，丙种鸟还会飞推测全世界所有的鸟都会飞。再比如，设2具有性质 $p$ ，4具有性质 $p$ ，6也具有性质 $p$ ，则由此推测所有的偶数都具有性质 $p$ 。这些问题都是归纳问题。对于一个归纳问题，由若干特殊事例所具有的特性推出全体事例所具有的特性，这种推理称为归纳推理。

归纳推理并不像演绎推理那样由前提必然推出结论，即由归纳得出的结论并不是百分之百对的，有可能出现错误。比如说，由三种鸟会飞推出所有鸟都会飞。所有的鸟都会飞吗？经过验证，并非如此！例如，鸵鸟和企鹅是鸟，但是鸵鸟和企鹅就不会飞。由此得出，上面通过归纳得出的结论是错误的。

科学发现中包含有许多归纳问题。事实上，一个科学结论的发现多半是通过归纳得到的。首先科学家验证几个特



例，然后由这几个特例归纳出一个一般结论。如果该结论后来被证明是对的，那么这就是一个科学发现；如果该结论后来被证明是错的，那么这就是一个归纳错误。归纳错误对于一个科学家来讲尽管是不情愿的，但是归纳错误是科学发现的基础，没有归纳错误就没有科学发现。

过去，归纳推理经常被描述成由前提以一定的程度得到结论，这种程度被定义为归纳强度。而归纳强度又由归纳概率来描述。归纳概率既不由前提本身决定，也不由结论本身决定，而由前提和结论之间的证据支持关系所决定。归纳概率事实上是一种概率的逻辑主义解释，由此而发展起的归纳逻辑的理论事实上是一种概率推理理论。但是，很显然，概率推理并不能概括所有的归纳问题。概率推理仅是归纳推理的一种特例。

利用Shafer信度理论可以给归纳推理以一般的描述。

所谓归纳就是在已知几种特殊情况成立以后问一般情况是否成立。这里，这几种特殊情况成立可以看成是确定一般情况是否成立的证据。

设我们用 $x$ 表示一般情况成立， $y$ 表示一般情况不成立，则可将集合 $\Theta = \{x, y\}$ 作为我们的识别框架。

一开始在我们知道几种特殊情况成立与一般情况是否成立之间无任何必然的联系。但经过我们的思考与分析，在我们的心目中可以产生一种信念，认为一般情况有多大可能成立。这就是说我们根据证据 $E$ 在框架 $\Theta$ 上产生了一个支持函数，即我们主观上认为已经知道的几种特殊情况的成立对结论成立具有多大的支持或在已知这几种特殊情况成立的条件下我们对结论成立具有多大的信度。

设证据 $E_1$ 以 $s_1$ 的程度支持 $\{x\}$ , 即我们认为在证据 $E_1$ 下一般情况成立的可能性为 $s_1$ , 则在证据 $E_1$ 下我们可以得到一个简单支持函数

$$S_1(A) = \begin{cases} s_1 & A = \{x\} \\ 1 & A = \Theta \\ 0 & \text{其他} \end{cases}$$

假设又知道另外的一些特殊情况成立, 则这些证据 $E_2$ 又可以在 $\Theta$ 上产生一个简单支持函数

$$S_2(A) = \begin{cases} s_2 & A = \{x\} \\ 1 & A = \Theta \\ 0 & \text{其他} \end{cases}$$

证据 $E_2$ 的得到对我们的信度有什么影响呢? 利用 Dempster 合成法则可求得 $S_1 \oplus S_2$ , 用来表示新证据的加入对我们的信度的影响。由前几章可知,  $S_1 \oplus S_2$ 将对 $\{x\}$ 提供 $s_1 + s_2 - s_1 \cdot s_2$ 的支持。这就是说, 证据 $E_2$ 使我们的信度增加了 $s_2 \cdot (1 - s_1)$ 。

假设又有一批证据 $E_3$ 支持 $\{x\}$ , 则同样可用 Dempster 合成法则求得 $E_3$ 对我们的信度的影响。

从以上分析中可以看出, 如果一开始一些特例使我们对结论成立的信度很大的话, 则再验证一个特例将使我们的信度增加很小。特别是, 当有若干特例使我们对结论成立的信度达到1时, 则不管再验证多少特例, 我们的信度都不会改变。由下面的论述可知, 验证一个反例可以使我们的信度改变很大(降到0)。

现在假设我们知道了一个反例, 即一批证据 $E_3$ 以1的程度支持 $\{y\}$ , 则证据 $E_3$ 与前面的证据是矛盾的。证据 $E_3$ 将使我们对 $\{x\}$ 的信度减到0, 即我们认为一般情况根本不可能

成立。

在这里，我们知道一个特例成立或不成立以后，对结论成立与否并不产生任何客观影响，但是据此我们对结论是否成立的信度可以增加或减少。所以知道某个特例成立或不成立只是意味着主观上信度的增加或减少，而没有任何客观的东西。

以上是我们用证据理论对归纳问题的粗略描述。由此可以看出，证据理论可以用于归纳问题的解决。

### §3 证据理论与非单调推理

传统逻辑推理满足单调性，即当假设增多时结论也随之增加。但是，日常生活中还存在有许多不满足单调性的逻辑推理。这些推理的共同特点是，公理集中的公理增加时，公理系统中的定理也可能增加，也可能减少。这样的逻辑推理称为非单调推理。

目前，非单调推理的方法基本上有两种<sup>[27]</sup>，一种是R. Reiter等提出的默认推理(default reasoning)，一种是J. McCarthy, D. McDermott等提出的界限推理(circumscription reasoning)。本节只涉及前一种。

默认推理是人们在知识不完全的情况下进行的推理。其基本的推理原则是：已知一般情况下 $A$ 是存在或成立的，如果现在没有证据可以证明 $A$ 不存在或不成立，那么就承认 $A$ 的存在或成立。以后若有进一步的事实证明 $A$ 不存在或不成立，那么我们认为 $A$ 存在或成立的信念可以修正。

例如，我们要判断一位大学教师张三是否掌握英语。一

般情况下，我们知道，大学教师都掌握英语。因此，根据默认推理，如果我们没有任何信息可以证明他不懂英语的话，那么我们就承认他懂英语。假设以后我们又得到了他不懂英语的信息，那么我们的信念可以加以修正，转而认为他不懂英语。

再比如，我们要判断某只鸟是否会飞。我们知道，大部分鸟是会飞的。因此，根据默认推理，如果我们没有任何关于这只鸟不会飞的证据，那么我们就默认这只鸟会飞。以后如果有新的证据证明这只鸟确实不会飞，那么我们可以修正我们认为这只鸟会飞的信念，重新认为它不会飞。

从这两个例子中可以看出，默认推理的一个基本出发点就是在没有任何证据可利用的情况下，默认一个结论成立。因为根据常识，该结论一般情况下或大多数情况下是成立的。这里常识起了巨大的作用，如果没有常识，那么就不能进行默认推理。另外，默认推理关于我们的信念的表述只有认为某结论成立和认为某结论不成立两种。但是，在实际情形下，我们可以出现不确定信念。比如说，我们可以有80%的把握认为它成立，10%的把握认为它不成立。默认推理不能表示这样一种信念。

按照Shafer证据理论，在默认推理中，“一般情况下结论 $A$ 成立”是我们判断 $A$ 成立与否的证据，“没有任何信息表明结论 $A$ 不成立”说明除了以上证据以外我们再没有其他证据。

设识别框架 $\Theta = \{A \text{ 成立}, A \text{ 不成立}\}$ ，证据“一般情况下结论 $A$ 成立”可以使我们对 $\{A \text{ 成立}\}$ 获得比较大的信任（尽管比较大，但也可以不是确信），设信任程度为 $r$ ，则证据

“一般情况下结论 $A$ 成立”在框架 $\Theta$ 上产生了一个简单支持函数 $S$ ， $S$ 满足

$$S(B) = \begin{cases} s & B = \{A \text{ 成立}\} \\ 1 & B = \Theta \\ 0 & \text{其他} \end{cases}$$

这里信任程度 $s$ 反映了我们对 $A$ 成立的信念，其大小因问题的不同而不同。比如说，我们要判断大学教师张三是否掌握英语，我们的识别框架可以设为 $\Theta = \{\text{张三掌握英语}, \text{张三不掌握英语}\}$ ，所用证据是“一般情况下大学教师都掌握英语”，该证据对 $\{\text{张三掌握英语}\}$ 可以提供比较大的支持，因为不掌握英语的大学教师是极其少见的。因此，我们对 $\{\text{张三掌握英语}\}$ 的信任程度 $s$ 就比较大，接近于1。但是，如果我们将张三局限在某所一般的大学中，那么证据“一般情况下，该所大学中的教师都掌握英语”对 $\{\text{张三掌握英语}\}$ 所提供的支持度 $s'$ 可能就没有 $s$ 大。其原因就是该所大学是一所一般的大学而不是著名大学或重点大学。

另外，由于我们不可能根据“一般情况下结论 $A$ 成立”完全推出 $A$ 成立，因此，证据“一般情况下结论 $A$ 成立”对 $\{A \text{ 成立}\}$ 所提供的支持度不可能严格等于1。所以由一般情况下结论 $A$ 成立就认为现在 $A$ 成立仅仅是一种默认而不是确信。

在默认推理中，如果我们根据证据“一般情况下结论 $A$ 成立”得到 $A$ 成立的默认以后，又有新的事实证明 $A$ 不成立，那么我们可以修正我们的信念认为 $A$ 不成立。利用 Dempster 合成法则，我们可以对这种信念的修正给出一种非常简单的描述。

设  $\Theta = \{A \text{ 成立}, A \text{ 不成立}\}$ ，证据“一般情况下结论  $A$  成立”在框架  $\Theta$  上产生了一个简单支持函数

$$S(B) = \begin{cases} s & B = \{A \text{ 成立}\} \\ 1 & B = \Theta \\ 0 & \text{其他} \end{cases}$$

“我们得到了新的事实证明  $A$  不成立”，说明我们又获得了新的证据，该证据严格支持  $\{A \text{ 不成立}\}$ ，即新的证据在  $\Theta$  上产生了另一个简单支持函数

$$S'(B) = \begin{cases} 1 & B = \{A \text{ 不成立}\} \\ 0 & \text{其他} \end{cases}$$

由 Dempster 合成法则， $S$  与  $S'$  的合成仍为  $S'$ ，因此合成证据对  $\{A \text{ 成立}\}$  提供的支持度为 0，对  $\{A \text{ 不成立}\}$  提供的支持度为 1。即新的事实（证据）的加入，使我们对  $A$  成立的信念降到 0，对  $A$  不成立的信念增到 1。即由认为  $A$  成立转到认为  $A$  不成立。

综上所述，默认推理实质上是一种特殊的似真推理，而证据理论可以作为另一种解决默认推理的方法。

## §4 识别证据

本节探讨证据理论在模式识别<sup>[25]</sup>中的应用。

### 一、模式识别问题

什么是模式呢？所谓模式指的是客观存在的特定的物体或现象。例如，某个人，某个手写字母，某台发动机的某个声音等都是模式的例子。

具有相同属性的模式的集合，称为模式类。例如，所有手写字母A构成一个模式类，所有手写字母B构成了另外一个模式类。模式识别指的是用机器模拟人对模式进行分类和描述。

模式可以用特征向量来表示。比如说，声音可以用它的语言特征向量来描述。模式也可用基元来表示。比如说，染色体可用组成它的几种基元表示成一个符号串。对应于模式的特征向量表示，传统的模式识别方法有统计决策法等；对应于模式的基元表示，传统的模式识别方法有句法结构识别法等。本节所讨论的模式仅限于能用特征向量表示的模式，能用基元表示的模式暂不涉及。

模式识别问题可以表述为：设 $F = \{x | x \text{ 是模式} \}$ 是一个模式集合，设该模式集合可以分为 $n$ 个类，分别用 $\omega_1, \omega_2, \dots, \omega_n$ 来表示，已知 $A_i = \{x_1^i, x_2^i, \dots, x_{n_i}^{(i)}\} \subset \omega_i (i=1, 2, \dots, n)$ ，问给定模式 $x$ 属于哪个类？

通俗地说就是：已知

$\omega_1$ 中的 $n(1)$ 个代表模式 $x_1^1, x_2^1, \dots, x_{n(1)}^{(1)}$

$\omega_2$ 中的 $n(2)$ 个代表模式 $x_1^2, x_2^2, \dots, x_{n(2)}^{(2)}$

.....

$\omega_n$ 中的 $n(n)$ 个代表模式 $x_1^n, x_2^n, \dots, x_{n(n)}^{(n)}$

据此判定给定的模式 $x$ 属于哪个模式类。

## 二、识别证据

在上面的模式识别问题中，判定模式 $x$ 属于哪个模式类，其依据就是知道 $\omega_1$ 中的 $n(1)$ 个代表模式， $\omega_2$ 中的 $n(2)$ 个代表模式， $\dots$ ， $\omega_n$ 中的 $n(n)$ 个代表模式。这些依据可以看成是我们识别 $x$ 属于哪个模式类的证据，称为识别证据。

识别证据的给出必须遵循 $n(i) \geq 1 (i=1, 2, \dots, n)$ ，即在 $\omega_1, \omega_2, \dots, \omega_n$ 的每一类中至少知道一个代表模式。

原因是，在已知 $F$ 可分为 $n$ 个类 $\omega_1, \omega_2, \dots, \omega_n$ 以后，要确切判定模式 $x$ 属于哪个类就必须指出 $\omega_1, \omega_2, \dots, \omega_n$ 中的全部元素。若确切指出了 $\omega_1$ 包含什么， $\omega_2$ 又包含什么， $\dots$ ， $\omega_n$ 又包含什么，那么我们就可以确切知道 $x$ 属于哪个类。使能确切判定 $x$ 属于哪个模式的证据称为全证据。可是，当样本很多时，就不可能指出 $\omega_1, \omega_2, \dots, \omega_n$ 中的全部元素，只能指出它们中的若干个代表，这若干个代表就确定了这一类模式的共同特性（有时尽管不能精确地确定）。在这种情况下，若有某一个 $n(i) = 0$ ，即对于 $\omega_i$ ，一个代表元素也没指出来，那么对这一类模式的特性就一点也不清楚了，从而不可能判定某个给定的模式是否属于该类。

所以说，对于一条识别证据，每个集合 $A_i$ 都不能是空集。

### 三、两个约定

✓ 设用 $\theta_i$ 代表 $x \in \omega_i (i=1, 2, \dots, n)$ ，则可用 $\Theta = \{\theta_1, \theta_2, \dots, \theta_n\}$ 作为我们的识别框架。识别证据的作用就是在框架 $\Theta$ 上产生一个似真函数 $pl$ 或支持函数 $S$ 。

我们的第一个约定是： $pl(\{\theta_i\})$ 的值和 $x$ 与 $A_i$ 中的点的相似程度成正比。

该约定来自于这样的直观理解：要想根据识别证据确定 $x$ 属于哪一类，关键是要对比 $x$ 与哪些 $A_i$ 中的元素更相似。比如说，若 $x$ 相似于 $A_1$ 中的元素，那么 $x \in \omega_1$ 的似真性就大；若 $x$ 与 $A_3$ 中的元素相似性差一点，那么 $x \in \omega_3$ 的似真性就小一



些。因此 $x \in \omega_i$ 的似真性取决于 $x$ 与 $A_i$ 中的元素的相似程度。若该程度大,那么 $x \in \omega_i$ 的似真性大;若该程度小,那么 $x \in \omega_i$ 的似真性就小。

我们的第二个约定是:识别证据是一致性证据。

该约定来自于下面的直观理解:要想使集合 $A = \{\theta'_1, \theta'_2, \dots, \theta'_p\}$ 获得比较大的支持,那么对 $\theta'_i (i=1, \dots, p)$ 的支持度都不应该低于某一水平。换句话说,若在一个集合中存在着似真度很低的元素,那么该集合不可能获得比较大的支持。

#### 四、相似程度

直观上,说两个模式 $x$ 与 $y$ 相似,就是说 $x$ 与 $y$ 的特征差不多。

设 $x$ 与 $y$ 所包含的特征有 $m$ 个,即:

$$x = (x_1, x_2, \dots, x_m)$$

$$y = (y_1, y_2, \dots, y_m)$$

则可用

$$\|x - y\| = \sqrt{(x_1 - y_1)^2 + (x_2 - y_2)^2 + \dots + (x_m - y_m)^2}$$

定义 $x$ 与 $y$ 的距离(欧氏距离)。

若 $\|x - y\|$ 比较小,说明 $x$ 与 $y$ 的各个特征相差不多;若 $\|x - y\|$ 比较大,说明 $x$ 与 $y$ 的各个特征相差较大。因此可用 $\|x - y\|$ 来描述 $x$ 与 $y$ 的相似程度。

如何来描述一个元素 $x$ 与一个集合 $A$ 中的各元素相似,或干脆说, $x$ 与集合 $A$ 相似呢?

设 $A = \{x_1, x_2, \dots, x_n\}$ , 每一个 $x_i$ 又都有 $m$ 个特征, 设为

$$x_i = (x_i^1, x_i^2, \dots, x_i^m) \quad (i=1, \dots, n)$$

$$\text{令 } y^j = \frac{1}{n} \sum_{i=1}^n x_i^j \quad (j=1, \dots, m)$$

则我们构造出一个元素  $y = (y^1, y^2, \dots, y^m)$ 。该元素的每一个特征是  $A$  中各元素对应特征的平均值。这个元素称为  $A$  的中心。

值得指出的是： $A$  的中心并不一定是  $A$  中的元素。它是由  $A$  中的元素构造出来的，可以作为  $A$  中元素的代表。

有了一个集合  $A$  的中心的概念以后，我们就可以定义一个元素  $x$  与一个集合  $A$  整体相似的概念了。

$x$  与  $A$  的距离定义为  $x$  与  $A$  的中心的距离，用  $\|x - A\|$  来表示。

$\|x - A\|$  可以反映  $x$  与  $A$  整体相似的程度。若  $\|x - A\|$  大，说明  $x$  与  $A$  的中心距离大，从而  $x$  与  $A$  的中心相似程度小，从整体上来讲， $x$  与  $A$  的相似程度也就小；若  $\|x - A\|$  小，说明  $x$  与  $A$  的中心距离小，从而  $x$  与  $A$  的中心相似程度大，从整体上来讲， $x$  与  $A$  的相似程度也就大。 $x$  与  $A$  的相似程度与  $x$  与  $A$  的距离  $\|x - A\|$  成反比。

### 五、识别证据下的支持函数

在上面的两个约定以及相似程度的定义下，可求到上面的识别证据在框架  $\Theta$  上产生的支持函数  $S$ 。

由第一个约定

$$pl(\{\theta_i\}) = \frac{C}{\|x - A_i\|} \quad (C \text{ 为常数})$$

$$(i = 1, \dots, n)$$

由第二个约定

$$\forall A \subset \Theta \quad \text{pl}(A) = \max_{\theta \in A} \text{pl}(\{\theta\}) = \max_{\theta \in A} \frac{C}{\|x - A_\theta\|}$$

$$= \frac{C}{\min_{\theta \in A} \|x - A_\theta\|}$$

( $A_\theta$  代表  $\theta$  对应的类的代表集)

当  $A = \Theta$  时

$$\text{pl}(\Theta) = \frac{C}{\min_{\theta \in \Theta} \|x - A_\theta\|} = \frac{C}{\min_{i \in \{1, \dots, n\}} \|x - A_i\|} = 1$$

$$\therefore C = \min_{i \in \{1, \dots, n\}} \|x - A_i\|$$

将  $C$  带回  $\text{pl}(A)$  的表达式中得

$$\forall A \subset \Theta \quad \text{pl}(A) = \frac{\min_{i \in \{1, \dots, n\}} \|x - A_i\|}{\min_{\theta \in A} \|x - A_\theta\|}$$

由此即得识别证据在框架  $\Theta$  上产生的支持函数  $S$

$$S(A) = \begin{cases} 1 - \frac{\min_{i \in \{1, \dots, n\}} \|x - A_i\|}{\min_{\theta \in A} \|x - A_\theta\|} & A \neq \phi \\ 0 & A = \phi \end{cases} \quad (13.4.1)$$

## 六、识别证据的权重

以上由识别证据得到的支持函数是一致支持函数，一致支持函数是可分离支持函数的特例，因此可以求到它所对应的证据的估价。

**定理13.4.1** 设  $\omega_1, \omega_2, \dots, \omega_n$  是  $n$  个模式类， $A_1, A_2, \dots, A_n$  分别是它们的代表集， $x$  是要识别的模式。又设识别框架  $\Theta = \{\theta_1, \theta_2, \dots, \theta_n\}$  ( $\theta_i$  表示  $x \in \omega_i$ ) 满足

$$\|x - A_1\| \leq \|x - A_2\| \leq \dots \leq \|x - A_n\|$$

则将  $A_1, A_2, \dots, A_n$  作为识别证据所得到的框架  $\Theta$  上的一致支持函数  $S_x$  所对应的证据的估价满足下列条件

$$(1) W_x(\{\theta_1, \dots, \theta_i\}) = \log \frac{\|x - A_{i+1}\|}{\|x - A_i\|} \quad (13.4.2)$$

$$i = 1, \dots, n-1$$

$$(2) W_x(A) = 0 \quad A \text{ 为其他 } \Theta \text{ 的真子集}$$

$$(3) W_x(\Theta) = \begin{cases} \infty & \text{如果(13.4.2)式对所有 } i \leq n-1 \text{ 都有限} \\ 0 & \text{其他} \end{cases}$$

证：根据证据的估价与可分离支持函数的一一对应性，只要证明满足(1)、(2)、(3)的估价  $W_x$  所对应的可分离支持函数  $S_x$  与(13.4.1)式定义的一致支持函数  $S_x$  一致即可！

由(1)、(2)、(3)可知，被  $W_x$  授予了正的权重的  $\Theta$  的子集是套形的，所以满足(1)、(2)、(3)的估价  $W_x$  所对应的可分离支持函数是一致的。

又由(5.4.6)式可知， $S_x$  对应的似真函数  $pl_x$  满足

$$\forall \theta_i \in \Theta$$

$$pl_x(\{\theta_i\}) = Q_x(\{\theta_i\}) = K \cdot \exp\left(-\sum_{\substack{B \subseteq \Theta \\ \theta_i \in B}} W_x(B)\right)$$

$$= K \cdot \exp\left(-\sum_{i=1}^{n-1} \log \frac{\|x - A_{i+1}\|}{\|x - A_i\|}\right)$$

$$= K \cdot \exp\left(-\log \frac{\|x - A_n\|}{\|x - A_1\|}\right)$$

$$= \frac{K \cdot \|x - A_i\|}{\|x - A_i\|}$$

所以由上小节的推导可知,  $S_x$  与(13.4.1)式中的  $S_x$  是一致的。 证毕

在定义 7.3.1 中我们针对证据的估价  $W: 2^\Theta \rightarrow [0, 1]$  将  $\sum_{\substack{\theta \in A \\ \theta' \notin A}} W(A)$  定义为有利于  $\theta$  而不利于  $\theta'$  的证据的权重。定理 13.4.1 中满足(1)、(2)、(3)的  $W_x$  是一个证据的估价, 因此对  $W_x$ , 我们也可以引入并求到有利于  $\theta$  而不利于  $\theta'$  的证据的权重。

**定理 13.4.2** 满足定理 13.4.1 中的(1)、(2)、(3)的证据的估价  $W_x$  满足

$$\sum_{\substack{\theta \in A \\ \theta' \notin A}} W_x(A) = \begin{cases} \log \frac{\|x - A_{\theta'}\|}{\|x - A_\theta\|} & \text{如果 } \|x - A_\theta\| < \|x - A_{\theta'}\| \\ 0 & \text{其他} \end{cases} \quad (13.4.3)$$

并且  $W_x$  是框架  $\Theta$  上满足(13.4.3)的唯一的—个估价。

证: 设  $\Theta = \{\theta_1, \dots, \theta_n\}$ , 其中  $\theta_1, \dots, \theta_n$  满足

$$\|x - A_1\| \leq \|x - A_2\| \leq \dots \leq \|x - A_n\|$$

(1) 先证  $W_x$  满足(13.4.3)式, 设

$$\theta = \theta_i, \quad \theta' = \theta_j,$$

$$\begin{aligned} \text{则} \quad \sum_{\substack{\theta \in A \\ \theta' \notin A}} W_x(A) &= \sum_{1 \leq k < j} W_x(\{\theta_1, \dots, \theta_k\}) \\ &= \sum_{1 \leq k < j} \log \frac{\|x - A_{k+1}\|}{\|x - A_k\|}. \end{aligned}$$

$$= \begin{cases} \log \frac{\|x - A_j\|}{\|x - A_i\|} & \text{如果 } i < j \\ 0 & \text{其他} \end{cases}$$

$$= \begin{cases} \log \frac{\|x - A_{\theta\theta}\|}{\|x - A_\theta\|} & \text{如果 } \|x - A_\theta\| < \|x - A_{\theta\theta}\| \\ 0 & \text{其他} \end{cases}$$

$$\left( \because \text{当 } \|x - A_\theta\| = \|x - A_{\theta\theta}\| \text{ 时, } \log \frac{\|x - A_{\theta\theta}\|}{\|x - A_\theta\|} = 0 \right)$$

(2) 再证唯一性。假设有一个估价  $W: 2^\Theta \rightarrow [0, \infty)$  满足 (13.4.3), 那么我们须证  $W = W_*$ 。

首先设  $1 \leq i < j \leq n$ , 且  $A$  是包含  $\theta_j$  但是不包含  $\theta_i$  的  $\Theta$  的子集。那么

$$W(A) \leq \sum_{\substack{\theta_j \in A \\ \theta_i \notin A}} W(B) = 0$$

因此对任何不是  $\{\theta_1, \dots, \theta_i\} (i \leq n-1)$  形式的  $\Theta$  的子集  $A$ ,  $W(A) = 0$ 。

现在, 设  $A = \{\theta_1, \dots, \theta_i\}$ ,  $1 \leq i \leq n-1$ , 则

$$W(\{\theta_1, \dots, \theta_i\}) = \sum_{\substack{\theta_{i+1} \in A \\ \theta_{i+1} \notin A}} W(A) = \log \frac{\|x - A_{i+1}\|}{\|x - A_i\|}$$

所以  $W$  满足 (13.4.2) 式。

最后由证据的估价的定义知: 如果对所有  $\Theta$  的真子集  $A$ ,  $W(A)$  都是有限的, 那么  $W(\Theta) = \infty$ , 否则  $W(\Theta) = 0$ 。

因此  $W = W_*$ 。 证毕

如果  $\Theta$  只有两个元素  $\theta$  和  $\theta'$  (即要识别的类只有两个),

并且满足  $\|x - A_\theta\| \leq \|x - A_{\theta'}\|$ , 那么(13.4.2)式退化为

$$W_x(\{\theta\}) = \sum_{\substack{\theta \in A \\ \theta' \notin A}} W_x(A) = \log \frac{\|x - A_{\theta'}\|}{\|x - A_\theta\|}$$

$$W_x(\{\theta'\}) = 0$$

因此, 估价  $W_x$  所决定的支持函数将退化为针对  $\{\theta\}$  的

权重为  $\log \frac{\|x - A_{\theta'}\|}{\|x - A_\theta\|}$  的简单支持函数。

## 七、进一步的探讨

得到支持函数  $S$  以后, 我们可以根据该支持函数进行识别。比如说, 我们可以取使  $pl(\{\theta\})$  最大的  $\theta_i$ , 将  $x$  归入  $\omega_i$  类。如果这样识别的话, 则实际上就是说  $x$  与哪个类的代表集整体上更相似, 则  $x$  就归入哪个类。这与人们直观的分类法是相符的。

当然也可能出现这样的情况, 使  $pl(\{\theta\})$  最大的  $\theta_i$  有好几个, 或有好几个  $\theta_i$ , 其  $pl(\{\theta\})$  都比较大, 以至于不能轻易删掉某一个。出现这种情况的话, 说明我们的识别证据还不够, 还必须寻找新的证据。

设又找到一批证据, 即又指出了  $\omega_1$  到  $\omega_n$  中的若干个代表模式。利用上面的方法又可得到框架  $\Theta$  上的一个支持函数。应用 Dempster 合成法则, 合成上面的两支持函数, 据此再进行  $x$  属于哪个类的判别。

当然, 得到该批证据以后, 我们也可以将该批证据合并到先前的证据中去, 依据上面的方法求得支持函数。

利用这两种方法得到的支持函数有什么区别呢?

前者由于利用了 Dempster 合成法则, 因此所得到的支持

函数将不再是一致的。后者得到的支持函数仍然是一致的。所以，第一种方法保持了证据间的冲突；第二种方法将最后的支持函数强加上一致性，隐藏了任何冲突的信息。但是，在模式识别问题中，后面得到的证据不应该与前面的证据有任何冲突，也即第二批证据与第一批证据应该总是一致的。因此得到第二批证据以后最好采取第二种方法计算支持函数。

不管利用哪一种方法，都说明我们的识别在出现错误或找不出类别的时候，可以补充证据进行修正，直到识别正确为止。这一点更加接近人类进行模式识别的方法。比如说，判断一个手写字母究竟是什么，首先找出若干个手写字母进行对比，若得不出结论，再找几个，直到得出结论为止。换句话说，我们这种模式识别的方法模拟了人类进行模式识别的过程。另外，这种方法也为知识的存贮提供了一种途径，因而也为机器学习提供了一种可能的方法。

## §5 专家系统中不确定性的处理

专家系统所用的基本技术就是产生式规则。产生式规则用来描述专家的经验，并且结合数据库得到新知识。产生式规则都是一些逻辑陈述，可以表为“IF 证据 THEN 结论”的形式。

可是专家的知识 and 经验经常带有不确定性，因而并不总是可以用“IF...THEN...”的形式来表示。比如说：“如果有A则通常有B”这种形式的知识用纯粹的“IF...THEN...”来表示显然是不合适的，因为由A并不一定能百分之百地推



出 $B$ 。

专家系统中的不确定性大致可以分为：

(1) 证据（前提）的不确定性。尽管由 $A$ 可推出 $B$ ，但是由于 $A$ 本身是不确定的，因而使得 $B$ 也带有不确定性。

(2) 规则的不确定性。此时尽管 $A$ 确定，可是由于由 $A$ 并不一定百分之百地推出 $B$ ，因而导致了结论 $B$ 的不确定。

(3) 证据与规则的不确定性。有时不仅证据不确定，规则也不确定，由此得出的结论当然不确定。

在目前的专家系统<sup>[26]</sup>中，这些不确定性都用 $[0, 1]$ 中的一个常数来表示。这个常数有的系统称为概率，有的系统称为必然性因子或可信度。比如说，证据的不确定性可用证据成立或出现的概率来表示；规则的不确定性由该规则成立的可信度来表示，等等。

下面列举两个专家系统，看一下在这些专家系统中不确定性到底是如何处理的。

PROSPECTOR是地质探矿专家系统。在PROSPECTOR中，证据的不确定性用该证据成立的概率或可信度来表示，多个证据的不确定性的处理采用乘积法则。首先给每个证据赋予一个可信度，当把这些证据组合起来作为条件时，总的可信度取决于各可信度的乘积。这从表面上来看符合概率理论，因此可将这种方法称为以概率为基础的方法。但是，这种方法事实上不能验证各分量的独立性，因此这是概率方法的不合法的应用。

PROSPECTOR对于规则的不确定性的处理采用函数的方法。这种方法建立在假设规则的不确定性由结论的可信度与证据的可信度之间的某一函数来描述的基础上。在这种方

法中，要确定函数形式，必须首先给出先验可信度。

在处理多个规则支持同一事实时的不确定性时，PROSPECTOR仍然采用概率的方法。首先引入一个可信性比例，把各个证据的可信性比例乘起来作为结论的可信性比例，然后由结论的可信性比例再转换为结论的可信度。

可见，PROSPECTOR处理不确定性是以条件概率和贝叶斯定理为基础的。我们也看到，PROSPECTOR对概率的解释是主观解释。因此在PROSPECTOR中有可信度和先验可信度的概念。

MYCIN是细菌感染医疗诊断咨询系统。由于在医生诊断疾病的过程中所用的信息并不适合于正规统计方法，因此采用可信度的概念进行推理。在MYCIN中，这种可信度所遵循的是MYCIN的设计者自己构造出来的规则，因此，MYCIN的设计者将这个可信度改称为必然性因子。

在MYCIN中引入两个数MB和MD，用于量度人们对于假设或断言的相信和不相信程度。

$MB[h, e] = x$ , 表示由于观察到证据 $e$ ，对假设 $h$ 的相信程度增加了 $x$ 。

$MD[h, e] = y$ , 表示由于观察到证据 $e$ ，对假设 $h$ 的相信程度减少了 $y$ 。

由于 $e$ 本身也可以是一个假设，因此可将MB和MD写为 $MB[h_1, h_2]$ ， $MD[h_1, h_2]$ 分别表示当 $h_2$ 成立时 $h_1$ 的相信程度的增加量和当 $h_2$ 成立时 $h_1$ 相信程度的减少量。

假设 $P(h)$ 是专家在不出现证据的条件下相信 $h$ 成立的先验概率。如果 $P(h|e) > P(h)$ ，则不相信程度的相对增加值为

$$\frac{P(h|e) - P(h)}{1 - P(h)}$$

以此值作为观察到证据 $e$ 以后,对假设 $h$ 相信程度增加的度量,称为 $MB[h, e]$ 。

同理当 $P(h|e) < P(h)$ 时,我们令

$$MD[h, e] = \frac{P(h) - P(h|e)}{P(h)}$$

用条件概率和先验概率可得到  $MB, MD$  的一般表示式

$$MB[h, e] = \begin{cases} 1 & \text{若 } P(h) = 1 \\ \frac{\max[P(h|e), P(h)] - P(h)}{\max[1, 0] - P(h)} & \text{其余} \end{cases}$$

$$MD[h, e] = \begin{cases} 1 & \text{若 } P(h) = 0 \\ \frac{\min[P(h|e), P(h)] - P(h)}{\min[1, 0] - P(h)} & \text{其余} \end{cases}$$

将  $MB$  和  $MD$  合为一个,可定义必然性因子  $CF$  如下

$$CF[h, e] = MB[h, e] - MD[h, e]$$

$CF$  可以取 $[-1, 1]$ 中的值。如果  $CF > 0$ , 表示系统相信假设成立; 如果  $CF < 0$ , 表示系统相信假设不成立。

当得到两个观察  $S_1$  和  $S_2$  时, 对一个假设的相信和不相信程度的量度可用下式计算

$$MB[h, s_1 \& s_2] = \begin{cases} 0 & \text{若 } MD[h, s_1 \& s_2] = 1 \\ MB[h, s_1] + MB[h, s_2] * (1 - MB[h, s_1]) & \text{其余} \end{cases}$$

$$MD[h, s_1 \& s_2] = \begin{cases} 0 & \text{若 } MB[h, s_1 \& s_2] = 1 \\ MD[h, s_1] + MD[h, s_2] * (1 - MD[h, s_1]) & \text{其余} \end{cases}$$

假设的合取和所取的可信度也可以从  $MB$  和  $MD$  的组合

来计算

$$MB[h_1 \& h_2, e] = \min(MB[h_1, e], MB[h_2, e])$$

$$MB[h_1 \text{ or } h_2, e] = \max(MB[h_1, e], MB[h_2, e])$$

用相似的公式计算 MD。

当证据不确定时

$$MB[h, s] = MB'[h, s] * \max(0, CF[s, e])$$

从以上式子可以看出可信度或必然性因子的合成法则无论是在形式上或内容上与 Dempster 合成法则都是相似的。可是 Dempster 合成法则要求，要合成的每项证据是独立的。在寻找规则与事实相匹配时，很难验证这样的独立性。因此这种 Dempster 合成法则的应用也是非法的。但是，当有些学者为概率陈述设计了另一种结构，即诊断树结构以后，这个问题得到了基本解决<sup>[3]</sup>。

不确定性可以分为随机性、模糊性和认识不确定性。

随机性在自然界中是大量存在的，因而建立专家系统时必然经常遇到。比如，某个地区有没有矿，这是随机的、不确定的。这种不确定性就是随机性。在 PROSPECTOR 中，这种随机性是通过由历史资料所得到的统计数字来描述的。例如，设在与本地区地质情况相类似的条件下，过去有几次有矿，有几次没矿，那么这一次在本地区有矿的概率就可以认为是过去有矿的次数与总次数的比。当然，如果没有这样的历史资料可利用的话，那么就只能依靠专家的主观估计了。研究随机性的理论是概率论和数理统计。

模糊性也是自然界和人类生活中大量存在的。比如说，“年青人”这个概念就是一个模糊概念，“很可能”也是一个模糊概念。“如果我是一个年青人，那么我很可能上得

去”则是一个模糊推理。在专家的经验中，包含有许多这种模糊概念和模糊推理。因此建造专家系统必须对这种不确定性作出处理。近年来发展起来的模糊集理论是处理模糊性的有力工具。这种理论在目前专家系统的建造中越来越受到人们的重视。

认识上的不确定性是由于人们认识水平的局限以及知识缺乏所造成的不确定性。如果说随机性和模糊性都是客观的不确定性的话，那么认识不确定性是主观的，认识上的。比如说，有两个命题 $A$ 和 $B$ ， $A$ 是否蕴含 $B$ 我不知道，但是根据我的知识，我相信 $A$ 蕴含 $B$ 成立的可能性是50%，即我对 $A \rightarrow B$ 成立的信任程度是50%。这种不确定性在人们的日常生活与推理中也是大量存在的。研究这种不确定性的理论过去有贝叶斯主观概率理论。事实上，现在所有基于概率的专家系统都以贝叶斯主观概率理论为基础。可是贝叶斯主观概率理论要求概率满足可加性。比如说，我们有以下规则：“如果启动器发出刺耳噪声，那么这是坏启动器的可能性是0.75。”按概率论，上述规则自动地意味着存在另一条以下规则：

“如果启动器发出刺耳噪声，那么它是好启动器的可能性是0.25。”但是，在许多场合，专家并不接受这样的规则。因此有必要寻求其他关于认识不确定性的理论。MYCIN的设计者们提出的关于必然性因子的理论是在这方面所迈出的一步，但是，它作为一种理论是相当不完善的。Shafer的证据理论也是关于认识不确定性的一种理论。这种理论目前也不是尽善尽美，但是与必然性因子的理论相比要完善的多，深刻的多。

Shafer的证据理论是研究不确定性的另一种理论。依据

这种理论研究问题，允许将整个大问题分解成许多小问题，将整个证据分解成一些小证据，在处理者对这些小证据和小问题作出处理以后，利用Dempster合成法则合成这些处理结果，从而得到整个问题的解答。在目前人工智能水平上，将大的分成小的的作法是必须的，因为只有这样，人们才能把握问题，否则无从下手。证据理论为这种分解与合成提供了理论依据，因此有可能制成一种专家系统，这种专家系统可以将问题和证据分成小的部分，又能将这些小问题和小证据联系起来进行处理。另外，建造这样的专家系统时，一开始所得到的可能仅仅是一个框架，以后随着系统的学习，可以更进一步丰富它的结构。

事实上，有资料表明<sup>[9]</sup>，Shafer和他所在的堪萨斯大学商学院人工智能研究组的同事们正在致力于研究这样一种系统。我们期待着他们能够获得成功。

在前面的第九章中，我们已经说过，专家的经验无论是确定的还是不确定的，都可用一个框架 $\Theta$ 上的基本可信度分配或信度函数来描述。这一点提醒我们去建造基于信度函数，而不是基于产生式规则的专家系统。事实上，本书第九章、第十章、第十一章、第十二章所论述的内容，都可以作为建造这样的专家系统的基础。比如说，基于第九章，可以建立专家集体决策系统或决策支持系统；基于第十章，可以建立决策分析专家系统；基于第十一章，可以建立预测专家系统；基于第十二章可建立指标体系确立专家系统，领导班子选取专家系统和医疗诊断专家系统等。期望不久的将来，这些类型的专家系统即可问世！

## 附录 本书用数学知识简介

### §1 集合及其运算

集合是数学中一个最原始的概念，不能确切定义，只能作一些说明。

所谓集合，是满足一定条件的若干个个体的汇集。这若干个个体称为集合的元素。集合通常用大写英文字母  $A, B, C, \dots$  或加一些合适的下标来表示，集合的元素通常用小写英文字母  $a, b, c, \dots$  或加一些合适的下标来表示。

集合的表示法基本上有两种。

一种是列举全部或部分元素的方法。用这种方法表示集合的一个首要前提是必须知道该集合的全部元素或它的元素的变化规律。将它的全部元素或有代表性的部分元素放到花括号中即可表示集合。例如，如果集合  $A$  包含有三个元素  $a, b, c$ ，则表示为  $A = \{a, b, c\}$ ；如果集合  $A$  包含有  $a_1, a_2, \dots, a_n$   $n$  个元素，则表示为  $A = \{a_1, a_2, \dots, a_n\}$ 。

另一种集合的表示方法是写出集合中的元素所满足的条件。例如，假设  $A$  是全体偶数的集合，则表示为

$$A = \{a \mid a \text{ 是偶数}\}$$

如果一个元素 $a$ 是某个集合 $A$ 的元素,则称为 $a$ 属于 $A$ (或 $A$ 包含元素 $a$ ),表示为 $a \in A$ 。如果元素 $a$ 不是某个集合 $A$ 的元素,则称为 $a$ 不属于 $A$ ,表示为 $a \notin A$ 。对于一个集合来讲,某一个元素是否属于该集合应该能够作出明确的判断,否则该集合不能称其为集合。

一个元素也不包含的集合称为空集,用 $\emptyset$ 来表示。

对于两个集合 $A$ 和 $B$ ,如果集合 $A$ 的元素都是集合 $B$ 的元素,即 $\forall a \in A$ 都有 $a \in B$ ( $\forall$ 表示任意或任取),则称集合 $A$ 是集合 $B$ 的子集,集合 $A$ 包含于或含于集合 $B$ ,集合 $B$ 包含集合 $A$ 或含集合 $A$ ,表示为 $A \subset B$ 或 $B \supset A$ 。如果集合 $A$ 不包含于集合 $B$ ,则表示为 $A \not\subset B$ 或 $B \not\supset A$ 。如果集合 $A$ 包含于 $B$ ,同时 $B$ 也包含于 $A$ ,即 $A \subset B$ , $B \subset A$ ,则称 $A$ 与 $B$ 相等,表示为 $A = B$ ;反之称为 $A$ 与 $B$ 不等,表示为 $A \neq B$ 。如果 $A \subset B$ 并且 $B \not\subset A$ ,则称 $A$ 是 $B$ 的真子集, $A$ 真包含于 $B$ , $B$ 真包含 $A$ ,表示为 $A \subset B$ 并且 $A \neq B$ ,或 $A \subset B$ , $A \neq B$ 。

我们规定空集 $\emptyset$ 包含于所有集合中,即对任意集合 $A$ , $\emptyset \subset A$ 。

一个集合也可以作为另外集合的元素,或者说一个集合的元素也可以是另外的集合。对于一个集合 $A$ ,我们称集合 $B = \{C | C \subset A\}$ 是集合 $A$ 的幂集,表示为 $2^A$ 。集合 $A$ 的幂集是 $A$ 的所有子集的集合。

如果 $A$ 是有限集,即 $A$ 的元素个数有限,则 $A$ 的元素的个数称为 $A$ 的基数,表示为 $|A|$ 。当 $|A| = n$ 时, $|2^A| = 2^n$ ,即当 $A$ 包含有 $n$ 个元素时, $A$ 的所有子集的个数为 $2^n$ 。这就是为什么将 $A$ 的幂集表示为 $2^A$ 的原因。

如果 $A$ 的基数是偶数,则 $(-1)^{|A|}$ 为 $+1$ ;如果 $A$ 的基



数是奇数, 则  $(-1)^{|A|}$  为  $-1$ 。 $(-1)^{|A|}$  称为  $A$  的 均 势 (parity)。另外, 若  $B \subset A$ , 则  $|A - B| = |A| - |B|$  而且  $(-1)^{|A-B|} = (-1)^{|A|} \cdot (-1)^{|B|}$ 。

如果  $A$  和  $B$  是两个集合, 那么我们称属于  $A$  并且属于  $B$  的元素构成的集合为  $A$  与  $B$  的交集, 属于  $A$  或者属于  $B$  的元素构成的集合为  $A$  与  $B$  的并集, 属于  $A$  并且不属于  $B$  的元素构成的集合为  $A$  与  $B$  的差集。分别记为  $A \cap B$ ,  $A \cup B$ ,  $A - B$ 。即

$$A \cap B = \{x \mid x \in A \text{ 并且 } x \in B\}$$

$$A \cup B = \{x \mid x \in A \text{ 或者 } x \in B\}$$

$$A - B = \{x \mid x \in A \text{ 并且 } x \notin B\}$$

特别是, 当我们的研究范围局限于某个特定的集合  $\Theta$  时,  $\Theta$  与  $\Theta$  的子集  $A$  的差集  $\Theta - A$  也表示为  $\bar{A}$ , 并称为集合  $A$  关于集  $\Theta$  的补集, 简称为  $A$  的补。

集合的并、交运算很容易推广到多个集合的情况。如果  $A_1, A_2, \dots, A_n$  是  $n$  个集合, 那么我们称同时属于各个集合的元素构成的集合为  $A_1, A_2, \dots, A_n$  的交, 记为  $A_1 \cap A_2 \cap \dots \cap A_n$  或  $\bigcap_{i \in \{1, 2, \dots, n\}} A_i$ 。如果  $A_1, A_2, \dots, A_n$  是  $n$  个集合, 那么我们称

属于  $A_1$  或者属于  $A_2$  ... 或者属于  $A_n$  的元素构成的集合为  $A_1, A_2, \dots, A_n$  的并, 记为  $A_1 \cup A_2 \cup \dots \cup A_n$  或  $\bigcup_{i \in \{1, 2, \dots, n\}} A_i$  ( $i = \{1, 2, \dots, n\}$ )。

由此即可定义并求到满足一定条件的若干个集合的并集或交集, 只要将条件写在  $\bigcup$  或  $\bigcap$  的下面即可! 例如, 设  $J = \{x \mid x \text{ 是偶数}\}$ , 则可定义并求到  $\bigcup_{i \in J} A_i$  和  $\bigcap_{i \in J} A_i$ 。如果有两个以上的

条件, 则可依次写在  $\bigcup$  或  $\bigcap$  的下面。例如,  $\bigcup_{\substack{i \in J \\ i \neq 0}} A_i$  表示对所

有 $A_i$  ( $i$ 是 $I$ 中不为0的元素) 求并。

## §2 等价关系和划分

假设 $A$ 和 $B$ 是两个集合,  $a \in A, b \in B$ , 则称 $\langle a, b \rangle$ 为一个序偶。序偶中的两个元素 $a, b$ 的位置不能交换, 否则将变为另外的序偶。另外当 $A$ 和 $B$ 是同一集合时, 序偶中的两个元素取自同一个集合。

对于集合 $A$ 和 $B$ , 我们称由所有序偶 $\langle a, b \rangle$  (其中 $a \in A, b \in B$ ) 构成的集合为集合 $A$ 与 $B$ 的笛卡尔乘积, 记为 $A \times B$ 。即

$$A \times B = \{ \langle a, b \rangle \mid a \in A, b \in B \}$$

集合 $A$ 与 $B$ 的笛卡尔乘积 $A \times B$ 的任意一个子集称为 $A$ 到 $B$ 的一个关系。关系通常记为 $R$ 或加一些合适的下标。例如, 设 $A = \{1, 2, 3\}, B = \{a, b\}$ , 则

$$A \times B = \{ \langle 1, a \rangle, \langle 1, b \rangle, \langle 2, a \rangle, \langle 2, b \rangle, \langle 3, a \rangle, \langle 3, b \rangle \}$$

$$\text{而 } R = \{ \langle 1, a \rangle, \langle 2, b \rangle, \langle 3, a \rangle \} \subset A \times B$$

就是从 $A$ 到 $B$ 的一个关系。

如果 $R$ 是从 $A$ 到 $A$ 的关系, 则称 $R$ 为 $A$ 上的或 $A$ 中的关系。

设 $R$ 是 $A$ 上的关系, 如果 $\forall a \in A, \langle a, a \rangle \in R$ , 则称 $R$ 是自返的或称 $R$ 满足自返性。

设 $R$ 是 $A$ 上的关系,  $\forall a \in A, \forall b \in A$ , 若由 $\langle a, b \rangle \in R$ 必然推出 $\langle b, a \rangle \in R$ , 则称 $R$ 是对称的或称 $R$ 满足对称性。

设 $R$ 是 $A$ 上的关系,  $\forall a \in A, \forall b \in A, \forall c \in A$ , 若由 $\langle a, b \rangle \in R$ 并且 $\langle b, c \rangle \in R$ 必然推出 $\langle a, c \rangle \in R$ , 则称 $R$ 是可传递的或称 $R$ 满足可传递性。

如果A上的关系R满足自返性、对称性以及可传递性,则称R为等价关系。若 $\langle a, b \rangle \in R$ , 则称a与b等价或a与b不可区分或a与b不可分辨。

设R是集合A上的等价关系。对于任何 $a \in A$ 来说, 定义集合 $[a]_R \subset A$ 为所有与a等价的元素所构成的集合, 即

$$[a]_R = \{b \mid b \in A \text{ 并且 } \langle a, b \rangle \in R\}$$

则称 $[a]_R$ 是由a所生成的R等价类。有时如不需要明确指出等价关系R, 则由a生成的等价类也可记为 $[a]$ 。

如果R是集合A上的等价关系, 则

(1) 对于所有的 $a, b \in A$ , 或者 $[a]_R = [b]_R$ 或者 $[a]_R \cap [b]_R = \phi$ ;

$$(2) \bigcup_{a \in A} [a]_R = A。$$

性质(1)说明, 只要 $b \in [a]_R$ , 则由b生成的R等价类与由a生成的R等价类就相等, 所以对于一个特定的等价类 $[a]_R$ , 就没有必要指明a究竟是什么, 只要指明它属于 $[a]_R$ 即可! 因此, 如果不需要明确指明等价类究竟是由哪个元素生成时, 该等价类也可以记为 $[\cdot]_R$ 或 $[\cdot]$ 。

性质(2)说明由a中所有元素生成的等价类必定覆盖整个集合A, 也即它们的并集是集合A。

给定非空集合A, 设非空集合 $B = \{B_1, B_2, \dots, B_m\}$ 满足:

$$(1) B_i \subset A, B_i \neq \phi \quad (i = 1, 2, \dots, m);$$

$$(2) \forall i, j \in \{1, 2, \dots, m\}, \text{ 如果 } i \neq j, \text{ 则 } B_i \cap B_j = \phi;$$

$$(3) \bigcup_{i=1}^m B_i = A;$$

则称 $B$ 是 $A$ 的一个划分。

值得指出的是：一般的集合 $A$ 的划分 $B$ 的元素个数可以有限也可以无限。具体到上定义也就是 $m$ 可以有限也可以无限。

由上面集合的划分的定义可知，如果 $R$ 是集合 $A$ 上的等价关系，则由 $A$ 中的元素生成的所有 $R$ 等价类的集合构成了 $A$ 的一个划分。这也就是说，由集合 $A$ 上的一个等价关系可以诱导出 $A$ 的一个划分。

事实上，该结论反过来也是对的，即由集合 $A$ 的一个划分也可以诱导出集合 $A$ 上的一个等价关系。

设 $B = \{B_1, B_2, \dots, B_m\}$ 是 $A$ 的一个划分，定义关系 $R$ 为： $A$ 中的两个元素 $a, b$ ，只要它们属于同一个 $B_i (i = 1, 2, \dots, m)$ ，则 $\langle a, b \rangle \in R$ 。可以验证，这样定义的关系 $R$ 是 $A$ 上的一个等价关系。

综上所述，等价关系与划分是一一对应的。

关于关系以及等价关系的更详细的讨论，请参看有关的离散数学书籍。

### §3 映射和函数

设 $f$ 是从 $A$ 到 $B$ 的关系，并且 $\forall a \in A$ 都存在唯一的一个 $b \in B$ 使得 $\langle a, b \rangle \in f$ ，则称 $f$ 为从 $A$ 到 $B$ 的映射或函数，记为 $f: A \rightarrow B$ 。特别是，当 $B$ 是某个实数集时， $f: A \rightarrow B$ 称为实函数，有时也简称为函数。设 $f: A \rightarrow B$ 是一个实函数， $A$ 是某些集合的集合，则称 $f$ 是从 $A$ 到 $B$ 的集函数。如果对于 $b \in B$ ，存在一个 $a \in A$ ，使得 $\langle a, b \rangle \in f$ ，则 $b$ 称为 $a$ 的函数值，记为 $b = f(a)$ 。

例如, 设  $\Theta = \{a, b, c, d\}$ , 定义  $p: \Theta \rightarrow [0, 1]$  为

$$\begin{aligned} p(a) &= 0.5 & p(b) &= 0.3 & p(c) &= 0.8 \\ p(d) &= 0.5 \end{aligned}$$

则  $p: \Theta \rightarrow [0, 1]$  为一实函数。

对上面的  $\Theta$ , 我们定义  $pl: 2^\Theta \rightarrow [0, 1]$  为

$$\begin{aligned} pl(\{a, b\}) &= 0.5 & pl(\{b, c\}) &= 0.3 \\ pl(\{a, b, d\}) &= 0.8 & pl(\{a, b, c, d\}) &= 0.5 \\ pl(A) &= 0 \quad (A \text{ 为其他 } \Theta \text{ 的子集}) \end{aligned}$$

则  $pl: 2^\Theta \rightarrow [0, 1]$  为一集函数。

对于函数  $f: A \rightarrow B$ , 如果对于任意的  $a_1, a_2 \in A$ ,  $a_1 \neq a_2$ , 都有  $f(a_1) \neq f(a_2)$ , 则称  $f$  是一对一的; 如果对于任意的  $b \in B$ , 至少存在一个  $a \in A$ , 使得  $f(a) = b$ , 则称  $f$  是到上的; 如果  $f$  既是一对一的, 又是到上的, 则称  $f$  是一一对应。

如果  $f: A \rightarrow B$  是一一对应, 则  $\forall b \in B$  都存在唯一的  $a \in A$  与之对应, 因此可定义一个从  $B$  到  $A$  的函数, 此函数称为  $f$  的反函数, 简称为  $f$  的反, 记为  $f^{-1}: B \rightarrow A$ 。若  $f(a) = b$ , 则  $f^{-1}(b) = a$ 。

例如, 指数函数  $f: \mathbb{R} \rightarrow \mathbb{R}^+$  定义为

$$\forall x \in \mathbb{R} \quad f(x) = e^x \quad (e^x \text{ 表示数 } e \text{ 的 } x \text{ 次方})$$

在本书中指数函数有时也用  $\exp(x)$  来表示, 即  $\exp(x) = e^x$ 。指数函数是一一对应, 因此可定义它的反函数, 指数函数的反函数是对数函数, 在本书中记为  $\log(x)$  (或记为  $\ln(x)$ )。

设  $f: A \rightarrow B$  和  $g: B \rightarrow C$  是两个函数, 定义函数  $g \circ f: A \rightarrow C$  满足

$$\forall a \in A \quad g \circ f(a) = g(f(a))$$

则函数  $g \circ f$  称为  $f$  和  $g$  的合成。

例如, 设  $A = \{1, 2, 3\}$ ,  $B = \{a, b\}$ ,  $C = \{p, q\}$ 。

$f: A \rightarrow B$  定义为

$$f = \{\langle 1, a \rangle, \langle 2, a \rangle, \langle 3, b \rangle\}.$$

$g: B \rightarrow C$  定义为

$$g = \{\langle a, p \rangle, \langle b, q \rangle\}.$$

则  $f$  与  $g$  的合成  $g \circ f: A \rightarrow C$  为

$$g \circ f = \{\langle 1, p \rangle, \langle 2, p \rangle, \langle 3, q \rangle\}$$

因为  $g \circ f(1) = g(f(1)) = g(a) = p$

$$g \circ f(2) = g(f(2)) = g(a) = p$$

$$g \circ f(3) = g(f(3)) = g(b) = q.$$

## §4 代数系统

代数系统是具有某些结构的集合, 因而也称为代数结构。

设  $A$  是一个集合, 我们用  $A^n$  表示  $A \times A \times \cdots \times A$ , 即  $n$  个  $A$  的笛卡尔乘积, 则从  $A^n$  到  $A$  的一个映射称为  $A$  上的 (或  $A$  中的) 一个  $n$  元运算。

设  $f: A^n \rightarrow A$  是  $A$  上的一个  $n$  元运算, 则  $\forall a_1, a_2, \dots, a_n \in A$  都存在唯一一个  $a \in A$  与之对应, 即对  $A$  中任意  $n$  个元素都可以通过运算  $f$  得到唯一的一个运算结果  $a$ , 可表示为  $a = a_1 f a_2 f \cdots f a_n$ 。以后运算经常用  $\oplus, \otimes, \circ, \cdot, *$  等来表示。

运算具有如下的一些性质:

1. 结合律: 设  $\oplus$  是  $A$  中的一个二元运算, 若  $\forall a, b, c \in A$

$$(a \oplus b) \oplus c = a \oplus (b \oplus c)$$

则称运算 $\oplus$ 满足结合律（或可结合性）。

2. 交换律：设 $\oplus$ 是A中的一个二元运算，若 $\forall a, b \in A$

$$a \oplus b = b \oplus a$$

则称运算 $\oplus$ 满足交换律。

3. 分配律：设 $\circ$ 和 $*$ 是A上的两个二元运算。若

$$\forall a, b, c \in A \quad a \circ (b * c) = (a \circ b) * (a \circ c)$$

则称运算 $\circ$ 对运算 $*$ 满足第一分配律。

若 $\forall a, b, c \in A \quad (b * c) \circ a = (b \circ a) * (c \circ a)$

则称运算 $\circ$ 对运算 $*$ 满足第二分配律。若运算 $\circ$ 对运算 $*$ 既满足第一分配律，又满足第二分配律，则称运算 $\circ$ 对运算 $*$ 满足分配律。

4. 幺元：设 $\oplus$ 为A上的二元运算，若 $\exists e \in A$ 使得 $\forall x \in A$

$$x \oplus e = e \oplus x = x$$

则称 $e$ 为该运算的一个幺元。

5. 零元：设 $\circ$ 为A上的二元运算，若 $\exists 0 \in A$ 使得 $\forall x \in A$

$$0 \circ x = x \circ 0 = 0$$

则称 $0$ 为该运算的一个零元。

6. 逆元：设 $\circ$ 为A上的二元运算， $e$ 为该运算的幺元，若对元素 $x$ ，存在元素 $x_i^{-1} \in A$ 使得

$$x_i^{-1} \circ x = e$$

则称 $x_i^{-1}$ 为 $x$ 对运算 $\circ$ 的左逆元素。若对A中元素 $x$ ，存在 $x_r^{-1} \in A$ 使得

$$x \circ x_r^{-1} = e$$

则称 $x_r^{-1}$ 是 $x$ 对运算 $\circ$ 的右逆元素。

一个运算的幺元若存在则必唯一。

一个运算的零元若存在则必唯一。

若运算  $*$  满足结合律, 则一个元素对于运算  $*$  的左右逆元必相等, 称之为该元素的逆元。逆元若存在也必唯一。

对任意一个非空集合  $A$ , 若在  $A$  中存在一些  $n$  元运算  $\oplus, \otimes, \dots, *$ , 则  $(A, \oplus, \otimes, \dots, *)$  称为一个代数系统。

若代数系统  $(A, \oplus)$  中的二元运算  $\oplus$  满足可结合性, 则称  $(A, \oplus)$  为半群。若半群  $(A, \oplus)$  中的运算  $\oplus$  还存在幺元, 则称  $(A, \oplus)$  为含幺半群; 若半群  $(A, \oplus)$  中的运算  $\oplus$  是可交换的, 则称  $(A, \oplus)$  为可交换半群; 若半群  $(A, \oplus)$  中的运算  $\oplus$  既满足可交换性, 又具有幺元, 则称  $(A, \oplus)$  为可交换含幺半群。

代数系统  $(A, \oplus)$  ( $\oplus$  为二元运算), 如果满足下列条件:

(1) 结合律: 即  $\forall a, b, c \in A$

$$(a \oplus b) \oplus c = a \oplus (b \oplus c)$$

(2) 存在幺元: 即  $\exists e \in A$  使得  $\forall x \in A$

$$x \oplus e = e \oplus x = x$$

(3) 存在逆元: 即  $\forall a \in A, \exists a^{-1} \in A$  使得

$$a \oplus a^{-1} = a^{-1} \oplus a = e$$

则称  $(A, \oplus)$  为群。若群  $(A, \oplus)$  中的二元运算  $\oplus$  还满足可交换性, 则称  $(A, \oplus)$  为可交换群或阿贝尔群。

代数系统  $(A, \oplus, \otimes)$ , 如果它的两个二元运算  $\oplus$  和  $\otimes$  满足下列条件:

(1)  $(A, \oplus)$  是一个可交换群;

(2)  $(A, \otimes)$  是一个半群;

(3) 运算  $\otimes$  对  $\oplus$  满足分配律。

则称  $(A, \oplus, \otimes)$  为环。

对于环  $(F, +, \cdot)$  (其中  $F$  至少包含一个以上的元素),



如果它的两个二元运算 $+$ 和 $\cdot$ 满足下列条件:

- (1)  $(F, \cdot)$ 有么元;
- (2)  $(F, \cdot)$ 是可交换的;
- (3)  $(F, \cdot)$ 中的任意非零元素都是可逆的。

则称 $(F, +, \cdot)$ 为域。

代数系统 $(B, +, \cdot, \neg)$ , 如果二元运算 $+$ 和 $\cdot$ 以及一元运算 $\neg$ 满足下列条件:

- (1) 交换律: 即 $\forall a, b \in B$   
$$a + b = b + a \quad a \cdot b = b \cdot a$$
- (2) 结合律: 即 $\forall a, b, c \in B$   
$$(a + b) + c = a + (b + c) \quad (a \cdot b) \cdot c = a \cdot (b \cdot c)$$
- (3) 分配律: 即 $\forall a, b, c \in B$   
$$a \cdot (b + c) = a \cdot b + a \cdot c \quad (b + c) \cdot a = b \cdot a + c \cdot a$$
  
$$a + (b \cdot c) = (a + b) \cdot (a + c)$$
  
$$(b \cdot c) + a = (b + a) \cdot (c + a)$$

(4) 存在么元: 对运算 $+$ 存在么元 $0$ , 对运算 $\cdot$ 存在么元 $1$ 。

- (5) 互补性: 即 $\forall a \in B$   
$$a + \bar{a} = 1 \quad a \cdot \bar{a} = 0$$

则称 $(B, +, \cdot, \neg)$ 为布尔代数。

在两个集合之间可定义映射的概念, 代数系统是定义了某种结构的特殊的集合, 因此同样可讨论两个代数系统之间的映射。

设 $(A, *)$ 和 $(B, \circ)$ 是两个代数系统,  $*$ 和 $\circ$ 是二元运算, 若映射 $g: A \rightarrow B$ 满足

$$\forall a_1, a_2 \in A \quad g(a_1 * a_2) = g(a_1) \circ g(a_2)$$

则称 $g$ 为从 $(A, *)$ 到 $(B, \circ)$ 的同态映射。若在 $(A, *)$ 和 $(B, \circ)$ 之间存在一个同态映射的话, 则称 $(A, *)$ 和 $(B, \circ)$ 同态。

如果从 $(A, *)$ 到 $(B, \circ)$ 的同态映射 $g: A \rightarrow B$ 是一一对应, 则称 $g$ 为从 $(A, *)$ 到 $(B, \circ)$ 的同构映射, 此时称 $(A, *)$ 和 $(B, \circ)$ 同构。

当代数系统 $(A, *)$ 和 $(B, \circ)$ 都是半群时, 相应的同态映射和同构映射就相应地称为半群同态映射和半群同构映射。同理, 还有群同态和群同构的概念。限于篇幅, 所有这些同态和同构的性质在此都不加以讨论, 有兴趣的读者可参看近世代数或离散数学方面的书籍。

## §5 数理逻辑

数理逻辑可以分为命题演算和谓词演算两大部分。

在命题演算中, 最基本的概念是命题。

一个陈述语句当它成立时我们称它为真, 当它不成立时我们称它为假。一个语句为真或假, 我们也称它取值为真或假。真或假有时也用1和0来表示。

一个陈述语句并不总是能判断出真或假的。对于能判断出真假的语句, 我们称为命题。因此, 命题是可以取值真或假, 并且只能取一个值的陈述语句。例如, 语句“北京是中国的首都”, 根据地理知识可知, 该语句成立, 因此它是一个命题, 取值为真或1。“禁止吸烟”是一个语句, 但它不能判断真假, 因此不是命题。

一般地讲, 命题可以分为两类, 一类是不能再分解为更简单命题的命题, 这种命题称为简单命题。另一类是可以分

解为若干更简单命题的命题，这种命题称为复合命题。例如：“北京是大城市并且是中国的首都”可以分解为“北京是大城市”和“北京是中国的首都”两个简单命题，因而是 一个复合命题。

命题通常用符号  $A, B, C, \dots$  等表示。一个命题  $A$  当它的真值不能具体给出（当  $A$  替换为任何具体的命题时，真值可以给出）时，称  $A$  为命题变元。因此要区分具体命题和命题变元。

通过引进联结词，我们可以由简单命题构造复合命题。

设  $A, B$  是命题，我们可以引进5种基本的逻辑联结词，非、合取、析取、蕴含、等价。两个命题经过逻辑联结词联结以后得到一复合命题，因此，可取唯一值1或0。

1. 非。非用符号  $\neg$  来表示，非  $A$  记作  $\neg A$  它表示命题  $A$  的否定，即与命题  $A$  对立的命题，取值情况如附表1所示。

2. 合取。合取用符号  $\wedge$  来表示。 $A$  合取  $B$  或  $A$  并且  $B$  记作  $A \wedge B$ 。 $A \wedge B$  表示这样的命题：当且仅当  $A$  和  $B$  都是真命题时，它才是真命题。具体的取值情况见附表2。

3. 析取。析取用符号  $\vee$  来表示。 $A$  析取  $B$  或  $A$  或者  $B$  记作  $A \vee B$ 。 $A \vee B$  表示这样的命题：当且仅当  $A$  和  $B$  中至少有一个是真命题时，它才是真命题。具体的取值情况见附表3。

4. 蕴含。蕴含有时也写为蕴涵。蕴含用符号  $\rightarrow$  来表示。 $A$  蕴含  $B$  记为  $A \rightarrow B$ 。 $A \rightarrow B$  有时也读作若  $A$  则  $B$ 。有时在  $A \rightarrow B$  中，我们也称  $A$  为前件， $B$  为后件。 $A \rightarrow B$  表示这样的命题：当且仅当  $A$  是真命题， $B$  是假命题时，它才是假命题。具体的取值情况见附表4。

5. 等价。等价用符号  $\sim$  来表示， $A$  等价于  $B$  可记为  $A \sim B$ 。

$A \sim B$  表示这样的命题：当且仅当  $A$  和  $B$  同时是真命题或者同时都是假命题时，它才是真命题。具体的取值情况见附表5。

附表 1

A	$\bar{A}$
0	1
1	0

附表 2

A	B	$A \wedge B$
0	0	0
0	1	0
1	0	0
1	1	1

附表 3

A	B	$A \vee B$
0	0	0
0	1	1
1	0	1
1	1	1

附表 4

A	B	$A \rightarrow B$
0	0	1
0	1	1
1	0	0
1	1	1

附表 5

A	B	$A \sim B$
0	0	1
0	1	0
1	0	0
1	1	1

有了上面定义的这5个逻辑连结词，我们可以从简单命题构造出复合命题。可是复合命题仍然是命题，因此在复合命题之间仍然可以使用连结词构造更复杂的复合命题。但是如何构造呢？当然不能随意，应有一定的规则。按照一定的规则构造出来的命题，我们称为合式公式。合式公式可以递归定义。这些规则是：

- (1) 一个命题变元是一个合式公式；
- (2) 若  $P$  是一个合式公式，则  $\bar{P}$  是一个合式公式；
- (3) 若  $P, Q$  是合式公式，则  $P \wedge Q, P \vee Q, P \rightarrow Q, P \sim Q$

都是合式公式；

(4) 合式公式只限于有限次地运用 (1), (2), (3) 所得到的符号串。

合式公式的取值可以通过真值表的方法来判定, 即将它的所有命题变元的取值组合列出来, 在这些不同的组合下, 看整个公式的取值如何? 例如, 我们上面列出的几个附表, 事实上就是对应于各个连结词的真值表。

一个合式公式如果在它的各个命题变元取值的不同组合下全都为真, 则称该公式为永真公式或重言式。如果对各个命题变元取值的不同组合, 取值全都为假, 则称为永假公式。不是永真公式的合式公式称为非永真公式, 不是永假公式的合式公式称为可满足公式。

可以证明, 一个合式公式的永真性, 永假性, 非永真性, 可满足性都是可判定的。即我们总能找到方法判断一个公式是不是永真公式, 是不是可满足公式等。

对于较复杂的合式公式, 在构造真值表时, 计算量可能会相当大, 构造过程也可能很冗长, 因此有必要寻找一种方法将复杂的公式化简为简单的公式。当然, 在化简的过程中要始终要求化简以后的公式与原来的公式是一回事或等值。

设  $\alpha, \beta$  是两个合式公式, 若对  $\alpha, \beta$  中的命题变元的各种不同组合,  $\alpha$  与  $\beta$  取值都是相同的, 则称公式  $\alpha$  与  $\beta$  等值。

在等值的前提下, 将复杂公式化简为简单公式的运算称为命题演算。一个合式公式可以最终化简为所有命题变元的析取的合取或合取的析取, 这样两种形式分别称为该合式公式的合取范式和析取范式。

设  $\alpha_1, \alpha_2, \dots, \alpha_n$  和  $\beta$  都是合式公式, 如果当  $\alpha_1 \wedge \alpha_2 \wedge$

$\cdots \wedge \alpha_n$  取值为1时,  $\beta$  也取值为1, 则称  $\beta$  是  $\alpha_1, \alpha_2, \dots, \alpha_n$  的逻辑结论 (推理),  $\alpha_1, \alpha_2, \dots, \alpha_n$  是  $\beta$  的逻辑前提, 也称  $\alpha_1, \alpha_2, \dots, \alpha_n$  可以推出  $\beta$ , 记为  $\alpha_1, \alpha_2, \dots, \alpha_n \Rightarrow \beta$ . 该定义是逻辑推理的一般定义。

例如, 由  $A \rightarrow B$  和  $B \rightarrow C$  可以推出  $A \rightarrow C$  就是著名的三段论推理。

以上简要介绍的是命题演算的基本内容, 关于谓词演算, 本书不打算再作介绍, 有兴趣的读者可参阅有关数理逻辑或离散数学方面的书籍。

## §6 关于 $\Sigma$ 的几个定理

**定理1** 如果  $A$  是一个有限集, 则

$$\sum_{B \subseteq A} (-1)^{|B|} = \begin{cases} 1 & A = \emptyset \\ 0 & A \neq \emptyset \end{cases}$$

证明: 根据二项式定理, 当  $n$  是一个正整数时, 有

$$\binom{n}{0} - \binom{n}{1} + \binom{n}{2} - \cdots + (-1)^n \binom{n}{n} = (1-1)^n = 0$$

设  $A = \{\theta_1, \dots, \theta_n\}$  是有限非空集合, 则

$$\begin{aligned} \sum_{B \subseteq A} (-1)^{|B|} &= (-1)^{|\emptyset|} + \sum_{i=1}^n (-1)^{|\{\theta_i\}|} \\ &\quad + \sum_{i < j} (-1)^{|\{\theta_i, \theta_j\}|} + \cdots + (-1)^{|A|} \\ &= \binom{n}{0} - \binom{n}{1} + \binom{n}{2} - \cdots + (-1)^n \binom{n}{n} \end{aligned}$$

$$= 0$$

设  $A = \emptyset$ , 则

$$\sum_{B \subseteq A} (-1)^{|B|} = (-1)^{|\emptyset|} = (-1)^0 = 1 \quad \text{证毕}$$

**定理2** 如果  $A$  是有限集,  $B \subseteq A$ , 则

$$\sum_{\substack{C \\ B \subseteq C \subseteq A}} (-1)^{|C|} = \begin{cases} (-1)^{|A|} & A = B \\ 0 & A \neq B \end{cases}$$

证明: 根据定理 1

$$\begin{aligned} \sum_{\substack{C \\ B \subseteq C \subseteq A}} (-1)^{|C|} &= \sum_{D \subseteq (A-B)} (-1)^{|B \cup D|} \\ &= (-1)^{|B|} \sum_{D \subseteq (A-B)} (-1)^{|D|} \\ &= \begin{cases} (-1)^{|B|} & A-B = \emptyset \\ 0 & A-B \neq \emptyset \end{cases} \\ &= \begin{cases} (-1)^{|A|} & A = B \\ 0 & A \neq B \end{cases} \quad \text{证毕} \end{aligned}$$

**定理3** 设  $\Theta$  是一个有限集,  $f$  和  $g$  都是定义在  $2^\Theta$  上的集函数, 则

$$\forall A \subseteq \Theta \quad f(A) = \sum_{B \subseteq A} g(B) \quad (6.1)$$

成立的充分必要条件为

$$\forall A \subseteq \Theta \quad g(A) = \sum_{B \subseteq A} (-1)^{|A-B|} f(B) \quad (6.2)$$

证明: (1) 先证(6.1)成立时(6.2)也成立。

假设(6.1)成立, 则由定理 2 知

$$\sum_{B \subseteq A} (-1)^{|A-B|} f(B) = (-1)^{|A|} \sum_{B \subseteq A} (-1)^{|B|} f(B)$$

$$\begin{aligned}
&= (-1)^{|A|} \sum_{B \subseteq A} (-1)^{|B|} \sum_{C \subseteq B} g(C) \\
&= (-1)^{|A|} \sum_{B \subseteq A} \sum_{C \subseteq B} (-1)^{|B|} g(C) \\
&= (-1)^{|A|} \sum_{C \subseteq A} \sum_{C \subseteq B \subseteq A} (-1)^{|B|} g(C) \\
&= (-1)^{|A|} \sum_{C \subseteq A} g(C) \sum_{C \subseteq B \subseteq A} (-1)^{|B|} \\
&= (-1)^{|A|} g(A) (-1)^{|A|} = g(A)
\end{aligned}$$

(2) 再证(6.2)成立时, (6.1)也成立。

假设(6.2)成立, 则由定理2知

$$\begin{aligned}
\sum_{B \subseteq A} g(B) &= \sum_{B \subseteq A} \sum_{C \subseteq B} (-1)^{|B-C|} f(C) \\
&= \sum_{C \subseteq A} \sum_{C \subseteq B \subseteq A} (-1)^{|B-C|} f(C) \\
&= \sum_{C \subseteq A} (-1)^{|C|} f(C) \sum_{C \subseteq B \subseteq A} (-1)^{|B|} \\
&= (-1)^{|A|} f(A) (-1)^{|A|} \\
&= f(A)
\end{aligned}$$

证毕

**定理4** 设  $\Theta$  是一个有限集,  $f$  和  $g$  都是定义在  $2^\Theta$  上的集函数, 则

$$\forall A \subseteq \Theta \quad f(A) = \sum_{B \subseteq A} (-1)^{|B|+1} g(B) \quad (6.3)$$

成立的充分必要条件为

$$\forall A \subseteq \Theta \quad g(A) = \sum_{B \subseteq A} (-1)^{|B|+1} f(B) \quad (6.4)$$

证明: (1) 先证(6.3)成立时, (6.4)也成立。

假设(6.3)成立; 则对  $\forall A \subseteq \Theta$



$$\begin{aligned}
f(A) &= \sum_{B \subset A} (-1)^{|B|+1} g(B) \\
\therefore \sum_{B \subset A} (-1)^{|B|+1} f(B) &= \sum_{B \subset A} (-1)^{|B|+1} \sum_{C \subset B} (-1)^{|C|+1} g(C) \\
&= \sum_{B \subset A} \sum_{C \subset B} (-1)^{|B|+1} (-1)^{|C|+1} g(C) \\
&= \sum_{C \subset A} \sum_{B: C \subset B \subset A} (-1)^{|B|} (-1)^{|C|} g(C) \\
&= \sum_{C \subset A} (-1)^{|C|} g(C) \sum_{C \subset B \subset A} (-1)^{|B|} \\
&= (-1)^{|A|} g(A) (-1)^{|A|} = g(A)
\end{aligned}$$

(2) 再证(6.4)成立时, (6.3)也成立。

将(6.3)中的  $f$  与  $g$  对调得(6.4)式, 将(6.4)式中的  $f$  与  $g$  对调又得(6.3)式, 这说明(6.3)与(6.4)式是对称的。所以由上面的(1)知: (6.4)成立时, (6.3)也成立 证毕

**定理5** 设  $\Theta$  是一个有限集,  $f$  和  $g$  都是定义在  $2^\Theta$  上的集函数, 则

$$\forall A \subset \Theta \quad f(A) = \sum_{B \subset A} (-1)^{|B|} g(B) \quad (6.5)$$

成立的充分必要条件为

$$\forall A \subset \Theta \quad g(A) = \sum_{B \subset A} (-1)^{|B|} f(\bar{B}) \quad (6.6)$$

证明:  $\forall A \subset \Theta$  令  $h(A) = -f(\bar{A})$ , 则  $h$  也是一个定义在  $2^\Theta$  上的集函数。

(1) 假设(6.5)成立, 则

$$\begin{aligned}
h(A) &= -f(\bar{A}) = - \sum_{B \subset \bar{A}} (-1)^{|B|} g(B) \\
&= \sum_{B \subset A} (-1)^{|B|+1} g(B) \quad (6.7)
\end{aligned}$$

由定理 4 知

$$\begin{aligned}\forall A \subset \Theta \quad g(A) &= \sum_{B \subset A} (-1)^{|B|+1} h(B) \\ &= \sum_{B \subset A} (-1)^{|B|} f(\bar{B})\end{aligned}$$

(2) 假设 (6.6) 成立, 则

$$g(A) = \sum_{B \subset A} (-1)^{|B|} f(\bar{B}) = \sum_{B \subset A} (-1)^{|B|+1} h(B)$$

由定理 4 知

$$\forall A \subset \Theta \quad h(A) = \sum_{B \subset A} (-1)^{|B|+1} g(B)$$

即

$$\forall A \subset \Theta \quad -f(\bar{A}) = \sum_{B \subset A} (-1)^{|B|+1} g(B)$$

$$\therefore f(A) = \sum_{B \subset A} (-1)^{|B|} g(B) \quad \text{证毕}$$

有了关于  $\Sigma$  的这几个定理以后, 本书第二章的大部分定理都能得到比较简单的证明, 有兴趣的读者不妨一试。

## 参 考 文 献

- [1] G. Shafer, *A Mathematical Theory of Evidence*, Princeton University Press, Princeton, 1976
- [2] G. Shafer, *Allocations of Probability*, Ann. Probab., 1979, V. 7, No. 5, 827—839
- [3] G. Shafer, *Probability Judgment in Artificial Intelligence and Expert Systems*, Working Paper No. 165, School of Business, The University of Kansas, Lawrence, 1984
- [4] G. Shafer, *The Construction of Probability Arguments*, Working Paper No. 183, 1986
- [5] G. Shafer, *Constructive Probability*, Synthese, 1981, V. 48, 1-60
- [6] G. Shafer, *Belief Function and Possibility Measure*, Working Paper No. 163, School of Business, The University of Kansas, Lawrence, 1984
- [7] P. P. Shenoy, G. Shafer, *Propagating Belief Functions With Local Computations*, Working Paper No. 184, School of Business, The University of Kansas, Lawrence, 1986

- ✓ [8] G. Shafer, R. Logan, *Implementing Dempster's Rule for Hierarchical Evidence*, Working Paper No. 174, School of Business, The University of Kansas, Lawrence, 1985
- ✓ [9] Z. Pawlak, *Rough Sets*, Internat. J. Computer. Inform. Sci., 1982, V. 11, No. 5, 341-356
- [10] P. Smets, *The Degree of Belief In A Fuzzy Event*, Information Science, 1981, V. 25, 1-19
- [11] P. Smets, *Medical diagnosis: Fuzzy Sets and Degrees of Belief*, Fuzzy Sets and Systems, 1981, V. 5, 259-266
- [12] D. Dubois, H. Prade, *On Several Representation of An Uncertain Body of Evidence*, In Fuzzy Information and Decision Processes (M. M. Gupta, E. Sanchez, eds.), North-Holland, 1982
- [13] D. Dubois, H. Prade, *A Set-theoretic View of Belief Function: Logical Operations and Approximations by Fuzzy Sets*, International Journal of General Systems, 1986, V. 12, 193-226
- [14] A. P. Dempster, *Upper and Lower Probabilities Induced by a Multivalued Mapping*, Annals of Mathematical Statistics, 1967, V. 38, 325-339
- [15] A. P. Dempster, *A Generalization of Bayesian Inference*, Journal of the Royal Statistical Society, 1968, Series B 30, 205-247
- [16] L. A. Zadeh, *On the Validity of Dempster's Rule of Combination of Evidence*, Memo No. UCB/ERL-M79/24, Univ. of California, Berkeley

- [17] L.A.Zadeh, *Fuzzy Set As a Basis for a Theory of Possibility*, Fuzzy Sets and Systems, 1978, V.1, 3—28
- [18] G.Choquet, *Theory of Capacities*, Ann.Inst.Fourier, Univ.(Grenoble)5 (1953—1954) 131—296
- [19] 陈淑敏:《集合论与数理逻辑初步》,黑龙江科技出版社1984年版
- [20] 徐洁磐:《离散数学导论》,高等教育出版社1982年版
- [21] 张连诚:《决策分析入门》,辽宁教育出版社1985年版
- [22] 李业:《预测学》,华南工学院出版社1986年版
- [23] 哈利·琼尼等,陆廷纲等译:《用于计划决策的技术预测》,复旦大学出版社1984年版
- [24] 伊·彼·苏斯洛夫,陆戈译:《统计指标理论》,黑龙江人民出版社1983年版
- [25] 许万雍:《模式识别原理》(一),清华大学计算机工程与科学系(讲义),1983年版
- [26] 傅京孙,蔡自兴,徐光祐:《人工智能及其应用》,清华大学出版社1988年版
- [27] 林尧瑞,张钹,石纯一:《专家系统原理与实践》,清华大学出版社1988年版